(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 745 676 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
04.12.1996 Bulletin 1996/49

(51) Int Cl.[6]: C12N 15/54, C12N 9/12,
C12Q 1/68

(21) Application number: 96303880.7

(22) Date of filing: 30.05.1996

(84) Designated Contracting States:
AT BE CH DE ES FI FR GB IT LI NL SE

(30) Priority: 31.05.1995 US 455686

(71) Applicant: Amersham Life Science Inc
Cleveland, Ohio 44128 (US)

(72) Inventors:
- Davis, Maria
Twinsburg, Ohio (US)
- Fuller, Carl
Cleveland Heights, Ohio 44118 (US)
- Moffett, Bruce
Shaker Heights, Ohio 44122 (US)

(74) Representative: Rollins, Anthony John
Group Patents,
Amersham International plc,
White Lion Road
Amersham, Bucks HP7 9LL (GB)

(54) Thermostable DNA polymerases

(57) An enzymatically active DNA polymerase having between 540 and 582 amino acids having a tyrosine at a position equivalent to position 667 of Taq DNA polymerase, wherein said polymerase lacks 5' to 3' exonuclease activity, and wherein said polymerase has at least 95% homology in its amino acid sequence to the DNA polymerase of Thermus aquaticus, Thermus flavus or Thermus thermophilus, and wherein said polymerase forms a single polypeptide band on an SDS PAGE.

EP 0 745 676 A1

Description

Background of the Invention

The present invention relates to novel thermostable DNA polymerases, the genes and vectors encoding them and their use in DNA sequencing.

US Patents 4,889,818 and 5,079,352 describe the isolation and expression of a DNA polymerase known as Taq DNA Polymerase (hereinafter referred to as Taq). It is reported that amino-terminal deletions wherein approximately one-third of the coding sequence is absent have resulted in producing a gene product that is quite active in polymerase assays. Taq is described as being of use in PCR (polymerase chain reaction).

US Patent No. 5,075,216 describes the use of Taq in DNA sequencing.

International patent application WO 92/06/06188 describes a DNA polymerase having an identical amino acid sequence to Taq except that it lacks the N-terminal 235 amino acids of Taq and its use in sequencing. This DNA polymerase is known as Δ Taq.

US Patent 4,795,699 describes the use of T7 type DNA polymerases (T7) in DNA sequencing. These are of great use in DNA sequencing in that they incorporate dideoxy nucleoside triphosphates (NTPs) with an efficiency comparable to the incorporation of deoxy NTPs; other polymerases incorporate dideoxy NTPs far less efficiently which requires comparatively large quantities of these to be present in sequencing reactions.

At the DOE Contractor-Grantee Workshop (Nov. 13-17, 1994, Santa Fe) and the I. Robert Lehman Symposium (Nov 11-14, 1994, Sonoma), Prof. S. Tabor identified a site in DNA polymerases that can be modified to incorporate dideoxy NTPs more efficiently. He reported that the presence or absence of a single hydroxy group (tyrosine vs. phenylalanine) at a highly conserved position on *E. coli,* DNA Polymerase I, T7, and Taq makes more than a 1000-fold difference in their ability to discriminate against dideoxy NTPs. (See also European Patent Application 94203433.1 published May 31, 1995, Publication No. 0 655 506 A1 and hereby incorporated by reference herein.)

Summary of the Invention

The present invention provides a DNA polymerase having an amino acid sequence differentiated from Taq in that it lacks the N-terminal 272 amino acids and has the phenylalanine at position 667 (of native Taq) replaced by tyrosine. Preferably, the DNA polymerase has methionine at position 1 (equivalent to position 272 of Taq) (hereinafter referred to as FY2) The full DNA sequence is given as Fig 1 (SEQ. ID. NO. 1). Included within the scope of the present invention are DNA polymerases having substantially identical amino acid sequences to the above which retain thermostability and efficient incorporation of dideoxy NTPs.

By a substantially identical amino acid sequence is meant a sequence which contains 540 to 582 amino acids that may have conservative amino acid changes compared with Taq which do not significantly influence thermostability or nucleotide incorporation, i.e. other than the phenylalanine to tyrosine conversion. Such changes include substitution of like charged amino acids for one another, or amino acids with small side chains for other small side chains, e.g., ala for val. More drastic changes may be introduced at noncritical regions where little or no effect on polymerase activity is observed by such a change.

The invention also features DNA polymerases that lack between 251 and 293 (preferably 271 or 272) of the N-terminal amino acids of *Thermus flavus* (Tfl) and have the phenylalanine at position 666 (of native Tfl) replaced by tyrosine; and those that lack between 253 and 295 (preferably 274) of the N-terminal amino acids of *Thermus thermophilus* (Tth) and have the phenylalanine at position 669 (of native Tth) replaced by tyrosine.

By efficient incorporation of dideoxy NTPs is meant the ability of a polymerase to show little, if any, discrimination in the incorporation of ddNTPs when compared with dNTPs. Suitably efficient discrimination is less than 1:10 and preferably less than 1:5. Such discrimination can be measured by procedures known in the art.

One preferred substantially identical amino acid sequence to that given above is that which contains 562 amino acids having methionine at position 1 and alanine at position 2 (corresponding to positions 271 and 272 of native Taq) (hereinafter referred to as FY3). A full DNA sequence is given as Fig. 2. This is a preferred DNA polymerase for expression by a gene of the present invention.

The purified DNA polymerases FY2 and FY3 both give a single polypeptide band on SDS polyacrylamide gels, unlike Δ Taq, having either a phenylalanine or tyrosine at position 667 which forms several polypeptide bands of similar size on SDS polyacrylamide gels.

A second preferred substantially identical amino acid sequence is that which lacks 274 of the N-terminal amino acids of Thermus thermophilus having methionine at position 1, and the phenylalanine to tyrosine mutation at position 396 (corresponding to position 669 of native Tth) (hereinafter referred to as FY4). A full DNA sequence is given as Fig. 5 (SEQ. ID. NO. 14).

The present invention also provides a gene encoding a DNA polymerase of the present invention. In order to assist

in the expression of the DNA polymerase activity, the modified gene preferably codes for a methionine residue at position 1 of the new DNA polymerase. In addition, in one preferred embodiment of the invention, the modified gene also codes for an alanine at position 2 (corresponding to position 272 of native Taq).

In a further aspect, the present invention provides a vector containing the gene encoding the DNA polymerase activity of the present invention, e.g., encoding an amino acid sequence differentiated from native Taq in that it lacks the N-terminal 272 amino acids and has phenylalanine at position 396 (equivalent to position 667 of Taq) replaced by tyrosine or a substantially identical amino acid sequence thereto.

In a yet further aspect, the present invention provides a host cell comprising a vector containing the gene encoding the DNA polymerase activity of the present invention, e.g., encoding an amino acid sequence differentiated from native Taq in that it lacks the N-terminal 272 amino acids and has phenylalanine at position 396 (equivalent to position 667 of native Taq) replaced by tyrosine or a substantially identical amino acid sequence thereto.

The DNA polymerases of the present invention are preferably in a purified form. By purified form is meant that the DNA polymerase is isolated from a majority of host cell proteins normally associated with it; preferably the polymerase is at least 10% (w/w) of the protein of a preparation, even more preferably it is provided as a homogeneous preparation, e.g., a homogeneous solution. Preferably the DNA polymerase is a single polypeptide on an SDS polyacrylamide gel.

The DNA polymerases of the present invention are suitably used in sequencing, preferably in combination with a pyrophosphatase. Accordingly, the present invention provides a composition which comprises a DNA polymerase of the present invention in combination with a pyrophosphatase, preferably a thermostable pyrophosphatase such as *Thermoplasma acidophilum pyrophosphatase*. (Schafer, G. and Richter, O.H. (1992) Eur. J. Biochem. 209, 351-355).

The DNA polymerases of the present invention can be constructed using standard techniques. By way of example, mutagenic PCR primers can be designed to incorporate the desired Phe to Tyr amino acid change (FY mutation) in one primer. In our hands these primers also carried restriction sites that are found internally in the sequence of the Taq polymerase gene clone of Delta Taq, pWB253, which was used by us as template DNA. However, the same PCR product can be generated with this primer pair from any clone of Taq or with genomic DNA isolated directly from *Thermus aquaticus*. The PCR product encoding only part of the gene is then digested with the appropriate restriction enzymes and used as a replacement sequence for the clone of Delta Taq digested with the same restriction enzymes. In our hands the resulting plasmid was designated pWB253Y. The presence of the mutation can be verified by DNA sequencing of the amplified region of the gene.

Further primers can be prepared that encode for a methionine residue at the N-terminus that is not found at the corresponding position of Taq, the sequence continuing with amino acid residue 273. These primers can be used with a suitable plasmid, e.g., pWB253Y DNA, as a template for amplification and the amplified gene inserted into a vector, e.g., pRE2, to create a gene, e.g., pRE273Y, encoding the polymerase (FY2). The entire gene can be verified by DNA sequencing.

Improved expression of the DNA polymerases of the present invention in the pRE2 expression vector was obtained by creating further genes, pREFY2pref (encoding a protein identical to FY2) and pREFY3 encoding FY3. A mutagenic PCR primer was used to introduce silent codon changes (i.e., the amino acid encoded is not changed) at the amino terminus of the protein which did not affect the sequence of the polypeptide. These changes led to increased production of FY2 polymerase. FY3 was designed to promote increased translation efficiency *in vivo*. In addition to the silent codon changes introduced in pREFY2pref, a GCT codon was added in the second position (SEQ. ID. NO. 2), as occurs frequently in strongly expressed genes in *E. coli*. This adds an amino acid to the sequence of FY2, and hence the protein was given its own designation FY3. Both constructs produce more enzyme than pRE273Y.

Silent codon changes such as the following increase protein production in *E. coli*:
substitution of the codon GAG for GAA;
substitution of the codon AGG, AGA, CGG or CGA for CGT *or* CGC;
substitution of the codon CTT, CTC, CTA, TTG or TTA for CTG; substitution of the codon ATA for ATT or ATC;
substitution of the codon GGG *or* GGA for GGT or GGC.

The present invention also provides a method for determining the nucleotide base sequence of a DNA molecule. The method includes providing a DNA molecule annealed with a primer molecule able to hybridize to the DNA molecule; and incubating the annealed molecules in a vessel containing at least one deoxynucleotide triphosphate, and a DNA polymerase of the present invention. Also provided is at least one DNA synthesis terminating agent which terminates DNA synthesis at a specific nucleotide base. The method further includes separating the DNA products of the incubating reaction according to size, whereby at least a part of the nucleotide base sequence of the DNA molecule can be determined.

In preferred embodiments, the sequencing is performed at a temperature above 50°C, 60°C, or 70°C.

In other preferred embodiments, the DNA polymerase has less than 1000, 250, 100, 50, 10 or even 2 units of exonuclease activity per mg of polymerase (measured by standard procedure, see below) and is able to utilize primers having only 4, 6 or 10 bases; and the concentration of all four deoxynucleoside triphosphates at the start of the incubating step is sufficient to allow DNA synthesis to continue until terminated by the agent, e.g., a ddNTP.

For cycle sequencing, the DNA polymerases of the present invention make it possible to use significantly lower amounts of dideoxynucleotides compared to naturally occurring enzymes. That is, the method involves providing an excess amount of deoxynucleotides to all four dideoxynucleotides in a cycle sequencing reaction, and performing the cycle sequencing reaction.

Preferably, more than 2, 5, 10 or even 100 fold excess of a dNTP is provided to the corresponding ddNTP.

In a related aspect, the invention features a kit or solution for DNA sequencing including a DNA polymerase of the present invention and a reagent necessary for the sequencing such as dITP, deaza GTP, a chain terminating agent such as a ddNTP, and a manganese-containing solution or powder and optionally a pyrophosphatase.

In another aspect, the invention features a method for providing a DNA polymerase of the present invention by providing a nucleic acid sequence encoding the modified DNA polymerase, expressing the nucleic acid within a host cell, and purifying the DNA polymerase from the host cell.

In another related aspect, the invention features a method for sequencing a strand of DNA essentially as described above with one or more (preferably 2, 3 or 4) deoxyribonucleoside triphosphates, a DNA polymerase of the present invention, and a first chain terminating agent. The DNA polymerase causes the primer to be elongated to form a first series of first DNA products differing in the length of the elongated primer, each first DNA product having a chain terminating agent at its elongated end, and the number of molecules of each first DNA products being approximately the same for substantially all DNA products differing in length by no more than 20 bases. The method also features providing a second chain terminating agent in the hybridized mixture at a concentration different from the first chain terminating agent, wherein the DNA polymerase causes production of a second series of second DNA products differing in the length of the elongated primer, with each second DNA product having the second chain terminating agent at its elongated end. The number of molecules of each second DNA product is approximately the same for substantially all second DNA products differing in length from each other by from 1 to 20 bases, and is distinctly different from the number of molecules of all the first DNA products having a length differing by no more than 20 bases from that of said second DNA products.

In preferred embodiments, three or four such chain terminating agents can be used to make different products and the sequence reaction is provided with a magnesium ion, or even a manganese or iron ion (e.g., at a concentration between 0.05 and 100 mM, preferably between 1 and 10 mM); and the DNA products are separated according to molecular weight in four or less lanes of a gel.

In another related aspect, the invention features a method for sequencing a nucleic acid by combining an oligonucleotide primer, a nucleic acid to be sequenced, between one and four deoxyribonucleoside triphosphates, a DNA polymerase of the present invention, and at least two chain terminating agents in different amounts, under conditions favoring extension of the oligonucleotide primer to form nucleic acid fragments complementary to the nucleic acid to be sequenced. For example, the chain terminating agent may be a dideoxynucleotide terminator for adenine, guanine, cytosine or thymine. The method further includes separating the nucleic acid fragments by size and determining the nucleic acid sequence. The agents are differentiated from each other by intensity of a label in the primer extension products.

While it is common to use gel electrophoresis to separate DNA products of a DNA sequencing reaction, those in the art will recognize that other methods may also be used. Thus, it is possible to detect each of the different fragments using procedures such as time of flight mass spectrometry, electron microscopy, and single molecule detection methods.

The invention also features an automated DNA sequencing apparatus having a reactor including reagents which provide at least two series of DNA products formed from a single primer and a DNA strand. Each DNA product of a series differs in molecular weight and has a chain terminating agent at one end. The reagents include a DNA polymerase of the present invention. The apparatus includes a separating means for separating the DNA product along one axis of the separator to form a series of bands. It also includes a band reading means for determining the position and intensity of each band after separation along the axis, and a computing means that determines the DNA sequence of the DNA strand solely from the position and intensity of the bands along the axis and not from the wavelength of emission of light from any label that may be present in the separating means.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

Description of the Preferred Embodiments

The drawings will first briefly be described.

Drawings

Figs 1-4 are the DNA sequences, and corresponding amino acid sequences, of FY2, FY3, and the DNA polymer-

ases of T. flavus and Thermus thermophilus, respectively. Figure 5 is the DNA sequence and corresponding amino acid sequence of FY4.

Examples

The following examples serve to illustrate the DNA polymerases of the present invention and their use in sequencing.

Preparation of FY DNA Polymerases (FY2 and FY3)

Bacterial Strains

E. coli strains: MV1190 [Δ(*srl - recA*) 306::Tn10, Δ *(lac-proAB), thi, supE,* F' *(traD36 proAB*$^+$ *lacI*^{q} *lacZ* ΔM15)]; DHλ$^+$ *[gyrA96, recA1, relA1, endA1, thi-1, hsdR17, supE44,* λ$^+$]; M5248 *[λ(bio275, cI857, cIII+, N+,* Δ *(H1))]*.

PCR

Reaction conditions based on the procedure of Barnes (91 Proc. Nat'l. Acad. Sci. 2216-2220, 1994) were as follows: 20mM Tricine pH8.8, 85mM KOAc, 200mM dNTPs, 10% glycerol, 5% DMSO, 0.5mM each primer, 1.5mM MgOAc, 2.5 U HotTub (Amersham Life Science Inc.) , 0.025 U DeepVent (New England Biolabs), 1-100 ng target DNA per 100ml reaction. Cycling conditions were 94°C 30s, 68°C 10m40s for 8 cycles; then 94°C 30s, 68°C 12m00s for 8 cycles; then 94°C 30s, 68°C13m20s for 8 cycles; then 94°C 30s, 68°C 14m40s for 8 cycles.

In vitro mutagenesis

Restriction enzyme digestions, plasmid preparations, and other *in vitro* manipulations of DNA were performed using standard protocols (Sambrook et al., Molecular Cloning 2nd Ed. Cold Spring Harbor Press, 1989). PCR (see protocol above) was used to introduce a Phe to Tyr amino acid change at codon 667 of native Taq DNA polymerase (which is codon 396 of FY2). Oligonucleotide primer 1 dGCTTGGGCAGAGGATCCGCCGGG (SEQ. ID. NO. 3) spans nucleotides 954 to 976 of the coding region of SEQ. ID. NO. 1 including a BamHI restriction site. Mutagenic oligo primer 2 dGGGATGGCTAGCTCCTGGGAGAGGCGGTGGCCGACATGCCGTAGA GGACCCCGTAGTTGATGG (SEQ. ID. NO. 4) spans nucleotides 1178 to 1241 including an NheI site and codon 396 of Sequence ID. NO. 1. A clone of exo-Taq deleted for the first 235 amino acids, pWB253 encoding DeltaTaq polymerase (Barnes, 112 Gene 29-35, 1992) was used as template DNA. Any clone of Taq polymerase or genomic DNA from *Thermus aquaticus* could also be utilized to amplify the identical PCR product. The PCR product was digested with BamHI and NheI, and this fragment was ligated to BamHI/NheI digested pWB253 plasmid to replace the corresponding fragment to create pWB253Y, encoding polymerase FY1. Cells of *E. coli* strain MV1190 were used for transformation and induction of protein expression, although any host strain carrying a *lac* repressor could be substituted. DNA sequencing verified the Phe to Tyr change in the coding region.

PCR primer 3 dGGAATTCCATATGGACGATCTGAAGCTCTCC (SEQ. ID. NO. 5) spanning the start codon and containing restriction enzyme sites, was used with PCR primer 4 dGGGGTACCAAGCTTCACTCCTTGGCGGAGAG (SEQ. ID. NO. 6) containing restriction sites and spanning the stop codon (codon 562 of Sequence ID. NO. 1). A methionine start codon and restriction enzyme recognition sequences were added to PCR primer 5 dGGAATTCCAT-ATGCTGGAGAGGCTTGAGTTT (SEQ. ID. NO. 7), which was used with primer 4 above. PCR was performed using the above primer pairs, and plasmid pWB253Y as template. The PCR products were digested with restriction enzymes NdeI and KpnI and ligated to NdeI/KpnI digested vector pRE2 (Reddi et al., 17 Nucleic Acids Research 10,473-10,488, 1989) to make plasmids pRE236Y, encoding FY1 polymerase, and pRE273Y encoding FY2 polymerase, respectively. Cells of *E. coli* strain DHλ$^+$ were used for primary transformation with this and all subsequent pRE2 constructions, and strain M5248 (λcI857) was used for protein expression, although any comparable pair of *E. coli* strains carrying the cI$^+$ and cI857 alleles could be utilized. Alternatively, any *rec*$^+$ *cI*$^+$ strain could be induced by chemical agents such as nalidixic acid to produce the polymerase. The sequences of both genes were verified. pRE273Y was found to produce a single polypeptide band on SDS polyacrylamide gels, unlike pRE253Y or pRE236Y.

Primer 6 dGGAATTCCATATGCTGGAACGTCTGGAGTTTGGCAGCCTC CTC (SEQ. ID. NO. 8) and primer 4 were used to make a PCR product introducing silent changes in codon usage of FY2. The product was digested with NdeI/BamHI and ligated to a pRE2 construct containing the 3' end of FY2 to create pREFY2pref, encoding FY2 DNA polymerase. Primer 7 dGGAATTCCATATGGCTCTGGAACGTCTGGAGTTTGGCAGCCTCCTC (SEQ. ID. NO. 9) and primer 4 were used to make a PCR product introducing an additional alanine codon commonly occurring at the second position of highly expressed genes. The NdeI/BamHI digested fragment was used as above to create pREFY3, encoding FY3

DNA polymerase.

Preparation of FY4 DNA Polymerase

Bacterial Strains

E. coli strains: DH1λ+ *[gyrA96, recA1, relA1, endA1, thi-1, hsdR17, supE44, λ+*]; M5248 [λ *(bio275, cI857, cIII+, N+,* Δ *(H1))*].

PCR

Genomic DNA was prepared by standard techniques from *Thermus thermophilus*. The DNA polymerase gene of *Thermus thermophilus* is known to reside on a 3 kilobase AlwNI fragment. To enrich for polymerase sequences in some PCR reactions, the genomic DNA was digested prior to PCR with AlwNI, and fragments of approximately 3 kb were selected by agarose gel electrophoresis to be used as template DNA. Reaction conditions were as follows: 10mM Tris pH8.3, 50mM KCl, 800μM dNTPs, 0.001% gelatin, 1.0μM each primer, 1.5mM $MgCl_2$, 2.5 U Tth, 0.025 U Deepvent (New England Biolabs), per 100μl reaction. Cycling conditions were 94°C 2 min, then 35 cycles of 94°C 30s, 55°C 30s, 72°C 3 min, followed by 72 °C for 7 min.

In vitro mutagenesis

Restriction enzyme digestions, plasmid preparations, and other *in vitro* manipulations of DNA were performed using standard protocols (Sambrook et al., 1989). Plasmid pMR1 was constructed to encode an exonuclease-free polymerase, with optimized codons for expression in *E. coli* at the 5' end. Primer 8 (SEQ. ID. NO. 10) (GGAATTCCAT-ATGCTGGAACGTCTGGAATTCGGCAGCCTC) was used with Primer 9 (SEQ. ID. NO.11) (GGGGTACCCTAACCCTT-GGCGGAAAGCCAGTC) to create a PCR product from *Tth* genomic DNA, which was digested with restriction enzymes NdeI and KpnI and inserted into plasmid pRE2 (Reddi et al., 1989, Nucleic Acids Research 17, 10473 - 10488) digested with the same enzymes.

To create the desired F396Y mutation, two PCR products were made from *Tth* chromosomal DNA. Primer 8 above was used in combination with Primer 10 (SEQ. ID. NO. 12) (GGGATGGCTAGCTCCTGGGAGAGCCTAT-GGGCGGACAT GCCGTAGAGGACGCCGTAGTTCACCG) to create a portion of the gene containing the F to Y amino acid change as well as a silent change to create an NheI restriction site. Primer 11 (SEQ. ID. NO. 13) (CTAGCTAGCCATCCCCTA CGAAGAAGCGGTGGCCT) was used in combination with primer 9 above to create a portion of the gene from the introduced NheI site to the stop codon at the 3' end of the coding sequence. The PCR product of Primers 8 and 10 was digested with NdeI and NheI, and the PCR product of Primers 9 and 11 was digested with NheI and KpnI. These were introduced into expression vector pRE2 which was digested with NdeI and KpnI to produce plasmid pMR5. In addition to the desired changes, pMR5 was found to have a spurious change introduced by PCR, which led to an amino acid substitution, K234R. Plasmid pMR8 was created to eliminate this substitution, by replacing the AflII/BamHI fragment of pMR5 for the corresponding fragment from pMR1. The FY4 polymerase encoded by plasmid pMR8 (SEQ. ID. NO. 14) is given in Figure 5.

Cells of *E. coli* strain DH1λ+ were used for primary transformation, and strain M5248 (λcI857) was used for protein expression, although any comparable pair of *E. coli* strains carrying the cI+ and cI857 alleles could be utilized. Alternatively, any *rec+ cI+* strain could be induced by chemical agents such as nalidixic acid to produce the polymerase.

Protein Sequencing

Determinations of amino terminal protein sequences were performed at the W.M. Keck Foundation, Biotechnology Resource Laboratory, New Haven, Connecticut.

Purification of Polymerases

A 1 liter culture of 2X LB (2% Bacto-Tryptone, 1% Bacto-Yeast Extract, 0.5% NaCl) + 0.2% Casamino Acids + 20 mM KPO_4 pH 7.5 + 50 μg/ml Ampicillin was inoculated with a glycerol stock of the appropriate cell strain and grown at 30°C with agitation until cells were in log phase (0.7-1.0 OD_{590}). 9 liters of 2X LB + 0.2% Casamino Acids + 20 mM KPO_4 pH 7.5 + 0.05% Mazu Anti-foam was inoculated with 1 liter of log phase cells in 10 liter Microferm Fermentors (New Brunswick Scientific Co.). Cells were grown at 30°C under 15 psi pressure, 350-450 rpm agitation, and an air flow rate of 14,000 cc/min ±1000 cc/min. When the OD_{590} reached 1.5-2.0, the cultures were induced by increasing the temperature to 40-42°C for 90-120 minutes. The cultures were then cooled to < 20°C and the cells harvested by

centrifugation in a Sorvall RC-3B centrifuge at 5000 rpm at 4°C for 15-20 minutes. Harvested cells were stored at -80°C.

Frozen cells were broken into small pieces and resuspended in pre-warmed (90-95°C) Lysis Buffer (20 mM Tris pH 8.5, 1 mM EDTA, 10 mM $MgCl_2$, 16 mM $(NH_4)_2SO_4$, 0.1% Tween 20, 0.1% Nonidet P-40, 1 mM PMSF). Resuspended cells were then heated rapidly to 80°C and incubated at 80°C for 20 minutes with constant stirring. The suspension was then rapidly cooled on ice. The cell debris was removed by centrifugation using a Sorvall GSA rotor at 10,000 rpm for 20 minutes at 4°C. The NaCl concentration of the supernatant was adjusted to 300 mM. The sample was then passed through a diethylaminoethyl cellulose (Whatman DE-52) column that had been previously equilibrated with Buffer A (20 mM Tris pH 8.5, 1 mM EDTA, 0.1% Tween 20, 0.1% Nonidet P-40, 300 mM NaCl, 10% glycerol, 1 mM DTT), and polymerase collected in the flow through. The sample was then diluted to a concentration of NaCl of 100mM and applied to a Heparin-sepharose column. The polymerase was eluted from the column with a NaCl gradient (100-500 mM NaCl). The sample was then dialyzed against Buffer B (20 mM Tris pH 8.5, 1 mM EDTA, 0.1% Tween 20, 0.1% Nonidet P-40, 10 mM KCl, 10% glycerol, 1 mM DTT) and further diluted as needed to lower the conductivity of the sample to the conductivity of Buffer B. The sample was then applied to a diethylaminoethyl (Waters DEAE 15 HR) column and eluted with a 10-500 mM KCl gradient. The polymerase was then diluted with an equal volume of Final Buffer (20 mM Tris pH 8.5, 0.1 mM EDTA, 0.5% Tween 20, 0.5% Nonidet P-40, 100 mM KCl, 50% glycerol, 1 mM DTT) and dialyzed against Final Buffer.

Assay of Exonuclease Activity

The exonuclease assay was performed by incubating 5 ul (25-150 units) of DNA polymerase with 5 ug of labelled [^{3}H]-pBR322 PCR fragment (1.6×10^4 cpm/ug DNA) in 100 ul of reaction buffer of 20 mM Tris·HCl pH 8.5, 5 mM $MgCl_2$, 10 mM KCl, for 1 hour at 60 °C. After this time interval, 200 ul of 1:1 ratio of 50 ug/ml salmon sperm DNA with 2 mM EDTA and 20% TCA with 2% sodium pyrophosphate were added into the assay aliquots. The aliquots were put on ice for 10 min and then centrifuged at 12,000g for 10 min. Acid-soluble radioactivity in 200 ul of the supernatant was quantitated by liquid scintillation counting. One unit of exonuclease activity was defined as the amount of enzyme that catalyzed the acid solubilization of 10 nmol of total nucleotide in 30 min at 60 °C.

Utility in DNA Sequencing

Example 1: DNA Sequencing with FY Polymerases (e.g., FY2 and FY3)

The following components were added to a microcentrifuge vial (0.5 ml) : 0.4 pmol M13 DNA (e.g., M13mp18, 1.0 µg); 2 µl Reaction Buffer (260 mM Tris-HCl, pH 9.5 65 mM $MgCl_2$); 2 µl of labeling nucleotide mixture (1.5 µM each of dGTP, dCTP and dTTP); 0.5 µl (5 µCi) of [a-^{33}P]dATP (about 2000Ci/mmol); 1 µl -40 primer (0.5 µM; 0.5 pmol/µl 5'GTTTTCCCAGTCACGAC-3'); 2 µl of a mixture containing 4 U/µl FY polymerase and 6.6 U/ml *Thermoplasma acidophilum* inorganic pyrophophatase (32 U/µl polymerase and 53 U/ml pyrophosphatase in 20 mM Tris (pH8.5), 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5% NP-40, 0.5% TWEEN-20 and 50% glycerol, diluted 8 fold in dilution buffer (10 mM Tris-HCl pH8.0, 1 mM 2-mercaptoethanol, 0.5% TWEEN-20, 0.5% NP-40)); and water to a total volume of 17.5 µl. These components (the labeling reaction) were mixed and the vial was placed in a constant-temperature water bath at 45°C for 5 minutes.

Four vials were labeled A, C, G, and T, and filled with 4 µl of the corresponding termination mix: ddA termination mix (150 µM each dATP, dCTP, dGTP, dTTP, 1.5 µM ddATP); ddT termination mix (150 µM each dATP, dCTP, dGTP, dTTP, 1.5 µM ddTTP); ddC termination mix (150 µM each dATP, dCTP, dGTP, dTTP, 1.5 µM ddCTP); ddG termination mix (150 µM each dATP, dCTP, dGTP, dTTP, 1.5 µM ddGTP).

The labeling reaction was divided equally among the four termination vials (4 µl to each termination reaction vial), and tightly capped.

The four vials were placed in a constant-temperature water bath at 72°C for 5 minutes. Then 4 µl of Stop Solution (95% Formamide 20 mM EDTA, 0.05% Bromophenol Blue, 0.05% Xylene Cyanol FF) added to each vial, and heated briefly to 70°-80°C immediately prior to loading on a sequencing gel (8% acrylamide, 8.3 M urea). Autoradiograms required an 18-36 hour exposure using Kodak XAR-5 film or Amersham Hyperfilm MP. High-quality sequence results with uniform band intensities were obtained. The band intensities were much more uniform than those obtained with similar protocols using *Taq* DNA polymerase or Δ*Taq* DNA polymerase.

Example 2: DNA Cycle Sequencing with FY Polymerases

The following components were added to a microcentrifuge vial (0.5 ml) which which is suitable for insertion into a thermocycler machine (e.g., Perkin-Elmer DNA Thermal Cycler): 0.05 pmol or more M13 DNA (e.g., M13mp18, 0.1 µg) , or 0.1 µg double-stranded plasmid DNA (e.g., pUC19); 2 µl Reaction Buffer (260 mM Tris-HCl, pH 9.5 65 mM

$MgCl_2$); 1 µl 3.0 µM dGTP; 1 µl 3.0 µM dTTP; 0.5 µl (5 µCi) of [α-^{33}P]dATP (about 2000Ci/mmol); 1 µl -40 primer (0.5 µM; 0.5 pmol/µl 5'GTTTTCCCAGTCACGAC-3'); 2 µl of a mixture containing 4 U/µl FY polymerase and 6.6 U/ml *Thermoplasma acidophilum* inorganic pyrophophatase (32 U/µl polymerase and 53 U/ml pyrophosphatase in 20 mM Tris (pH8.5), 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5% NP-40, 0.5% TWEEN-20 and 50% glycerol, diluted 8 fold in dilution buffer (10 mM Tris-HCl pH8.0, 1 mM 2-mercaptoethanol, 0.5% TWEEN-20, 0.5% NP-40)); and water to a total volume of 17.5 µl.

These components (labeling reaction mixture) were mixed and overlaid with 10 µl light mineral oil (Amersham). The vial was placed in the thermocycler and 30-100 cycles (more than 60 cycles is unnecessary) from 45°C for 1 minute to 95°C for 0.5 minute performed. (Temperatures can be cycled from 55°-95°C, if desired). The temperatures may be adjusted if the melting temperature of the primer/template is significantly higher or lower, but these temperatures work well for most primer-templates combinations. This step can be completed in about 3 minutes per cycle.

Four vials were labeled A, C, G, and T, and filled with 4 ml of the corresponding termination mix: ddA termination mix (150 µM each dATP, dCTP, dGTP, dTTP, 1.5 µM ddATP); ddT termination mix (150 µM each dATP, dCTP, dGTP, dTTP, 1.5 µM ddTTP); ddC termination mix (150 µM each dATP, dCTP, dGTP, dTTP, 1.5 µM ddCTP); ddG termination mix (150 µM each dATP, dCTP, dGTP, dTTP, 1.5 µM ddGTP). No additional enzyme is added to the termination vials. The enzyme carried in from the prior (labeling) step is sufficient.

The cycled labeling reaction mixture was divided equally among the four termination vials (4 µl to each termination reaction vial), and overlaid with 10 µl of light mineral oil.

The four vials were placed in the thermocycler and 30-200 cycles (more than 60 cycles is unnecessary) performed from 95°C for 15 seconds, 55°C for 30 seconds, and 72°C for 120 seconds. This step was conveniently completed overnight. Other times and temperatures are also effective.

Six µl of reaction mixture was removed (avoiding oil), 3 µl of Stop Solution (95% Formamide 20 mM EDTA, 0.05% Bromophenol Blue, 0.05% Xylene Cyanol FF) added, and heated briefly to 70°-80°C immediately prior to loading on a sequencing gel. Autoradiograms required an 18-36 hour exposure using Kodak XAR-5 film or Amersham Hyperfilm MP. High-quality sequence results with uniform band intensities were obtained. The band intensities were much more uniform than those obtained with similar protocols using *Taq* DNA polymerase or Δ*Taq* DNA polymerase.

Example 3: Sequencing with dGTP analogs to eliminate compression artifacts.

For either of the sequencing methods outlined in examples 1 and 2, 7-Deaza-2'deoxy-GTP can be substituted for dGTP in the labeling and termination mixtures at exactly the same concentration as dGTP. When this substitution is made, secondary structures on the gels are greatly reduced. Similarly, 2'-deoxyinosinetriphosphate can also be substituted for dGTP but its concentration must be 10-fold higher than the corresponding concentration of dGTP. Substitution of dITP for dGTP is even more effective in eliminating compression artifacts than 7-deaza-dGTP.

Example 4: Other Sequencing methods using FY polymerases

FY polymerases have been adapted for use with many other sequencing methods, including the use of fluorescent primers and fluorescent-dideoxy-terminators for sequencing with the ABI 373A DNA sequencing instrument.

Example 5: SDS-Polyacrylamide Gel Electrophoresis

Protein samples were run on a 14 X 16 mm 7.5 or 10% polyacrylamide gel. (Gels were predominantly 10% Polyacrylamide using a 14 X 16 mm Hoefer apparatus. Other sizes, apparatuses, and percentage gels are acceptable. Similar results can also be obtained using the Pharmacia Phast Gel system with SDS, 8-25% gradient gels. Reagent grade and ultrapure grade reagents were used.) The stacking gel consisted of 4% acrylamide (30:0.8, acrylamide: bisacrylamide), 125 mM Tris-HCl pH 6.8, 0.1% Sodium Dodecyl Sulfate (SDS). The resolving gel consisted of 7.5 or 10% acrylamide (30:0.8, acrylamide: bisacrylamide), 375 mM Tris-HCl pH 8.8, 0.1% SDS. Running Buffer consisted of 25 mM Tris, 192 mM Glycine and 0.1% SDS. 1X Sample Buffer consisted of 25 mM Tris-HCl pH 6.8, 0.25% SDS, 10% Glycerol, 0.1M Dithiothreitol, 0.1% Bromophenol Blue, and 1mM EDTA. A 1/4 volume of 5X Sample Buffer was added to each sample. Samples were heated in sample buffer to 90-100°C for approximately 5 minutes prior to loading. A 1.5 mm thick gel was run at 50-100 mA constant current for 1-3 hours (until bromophenol.blue was close to the bottom of the gel). The gel was stained with 0.025% Coomassie Blue R250 in 50% methanol, 10% acetic acid and destained in 5% methanol, 7% acetic acid solution. A record of the gel was made by taking a photograph of the gel, by drying the gel between cellulose film sheets, or by drying the gel onto filter paper under a vacuum.

Other embodiments are within the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: AMERSHAM LIFE SCIENCE

(ii) TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES

(iii) NUMBER OF SEQUENCES: 14

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Lyon & Lyon
(B) STREET: 633 West Fifth Street Suite 4700
(C) CITY: Los Angeles
(D) STATE: California
(E) COUNTRY: U.S.A.
(F) ZIP: 90071-2066

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: IBM P.C. DOS 5.0
(D) SOFTWARE: Word Perfect 5.1

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: To Be Assigned
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

Prior applications total, including application described below: one

(A) APPLICATION NUMBER: US 08/455,686
(B) FILING DATE: May 31, 1995

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Warburg, Richard J.
(B) REGISTRATION NUMBER: 32,327
(C) REFERENCE/DOCKET NUMBER: 219/304-PCT

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (213) 489-1600
(B) TELEFAX: (213) 955-0440
(C) TELEX: 67-3510

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1686 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: FY2
(B) LOCATION: 1...1683

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
ATG CTG GAG AGG CTT GAG TTT GGC AGC CTC CTC CAC GAG TTC GGC CTT    48
Met Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu
 1               5                  10                  15

CTG GAA AGC CCC AAG GCC CTG GAG GAG GCC CCC TGG CCC CCG CCG GAA    96
Leu Glu Ser Pro Lys Ala Leu Glu Glu Ala Pro Trp Pro Pro Pro Glu
            20                  25                  30

GGG GCC TTC GTG GGC TTT GTG CTT TCC CGC AAG GAG CCC ATG TGG GCC   144
```

Gly Ala Phe Val Gly Phe Val Leu Ser Arg Lys Glu Pro Met Trp Ala
35 40 45

GAT CTT CTG GCC CTG GCC GCC GCC AGG GGG GGC CGG GTC CAC CGG GCC 192
Asp Leu Leu Ala Leu Ala Ala Ala Arg Gly Gly Arg Val His Arg Ala
50 55 60

CCC GAG CCT TAT AAA GCC CTC AGG GAC CTG AAG GAG GCG CGG GGG CTT 240
Pro Glu Pro Tyr Lys Ala Leu Arg Asp Leu Lys Glu Ala Arg Gly Leu
65 70 75 80

CTC GCC AAA GAC CTG AGC GTT CTG GCC CTG AGG GAA GGC CTT GGC CTC 288
Leu Ala Lys Asp Leu Ser Val Leu Ala Leu Arg Glu Gly Leu Gly Leu
85 90 95

CCG CCC GGC GAC GAC CCC ATG CTC CTC GCC TAC CTC CTG GAC CCT TCC 336
Pro Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser
100 105 110

AAC ACC ACC CCC GAG GGG GTG GCC CGG CGC TAC GGC GGG GAG TGG ACG 384
Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr
115 120 125

GAG GAG GCG GGG GAG CGG GCC GCC CTT TCC GAG AGG CTC TTC GCC AAC 432
Glu Glu Ala Gly Glu Arg Ala Ala Leu Ser Glu Arg Leu Phe Ala Asn
130 135 140

CTG TGG GGG AGG CTT GAG GGG GAG GAG AGG CTC CTT TGG CTT TAC CGG 480
Leu Trp Gly Arg Leu Glu Gly Glu Glu Arg Leu Leu Trp Leu Tyr Arg
145 150 155 160

GAG GTG GAG AGG CCC CTT TCC GCT GTC CTG GCC CAC ATG GAG GCC ACG 528
Glu Val Glu Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala Thr
165 170 175

GGG GTG CGC CTG GAC GTG GCC TAT CTC AGG GCC TTG TCC CTG GAG GTG 576
Gly Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val
180 185 190

GCC GAG GAG ATC GCC CGC CTC GAG GCC GAG GTC TTC CGC CTG GCC GGC 624
Ala Glu Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly
195 200 205

CAC CCC, TTC AAC CTC AAC TCC CGG GAC CAG CTG GAA AGG GTC CTC TTT 672
His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe
210 215 220

GAC GAG CTA GGG CTT CCC GCC ATC GGC AAG ACG GAG AAG ACC GGC AAG 720
Asp Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys
225 230 235 240

CGC TCC ACC AGC GCC GCC GTC CTG GAG GCC CTC CGC GAG GCC CAC CCC 768
Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro
245 250 255

ATC GTG GAG AAG ATC CTG CAG TAC CGG GAG CTC ACC AAG CTG AAG AGC 816
Ile Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser
260 265 270

ACC TAC ATT GAC CCC TTG CCG GAC CTC ATC CAC CCC AGG ACG GGC CGC 864
Thr Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg
275 280 285

CTC CAC ACC CGC TTC AAC CAG ACG GCC ACG GCC ACG GGC AGG CTA AGT 912
Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser
290 295 300

AGC TCC GAT CCC AAC CTC CAG AAC ATC CCC GTC CGC ACC CCG CTT GGG 960
Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly
305 310 315 320

CAG AGG ATC CGC CGG GCC TTC ATC GCC GAG GAG GGG TGG CTA TTG GTG 1008
Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val
325 330 335

GCC CTG GAC TAT AGC CAG ATA GAG CTC AGG GTG CTG GCC CAC CTC TCC 1056
Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser
340 345 350

GGC GAC GAG AAC CTG ATC CGG GTC TTC CAG GAG GGG CGG GAC ATC CAC 1104
Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His
355 360 365

ACG GAG ACC GCC AGC TGG ATG TTC GGC GTC CCC CGG GAG GCC GTG GAC 1152
Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp
370 375 380

CCC CTG ATG CGC CGG GCG GCC AAG ACC ATC AAC TAC GGG GTC CTC TAC 1200
Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Tyr Gly Val Leu Tyr
385 390 395 400

GGC ATG TCG GCC CAC CGC CTC TCC CAG GAG CTA GCC ATC CCT TAC GAG 1248
Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu
405 410 415

GAG GCC CAG GCC TTC ATT GAG CGC TAC TTT CAG AGC TTC CCC AAG GTG 1296
Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val
420 425 430

CGG GCC TGG ATT GAG AAG ACC CTG GAG GAG GGC AGG AGG CGG GGG TAC 1344
Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr
435 440 445

```
GTG GAG ACC CTC TTC GGC CGC CGC CGC TAC GTG CCA GAC CTA GAG GCC   1392
Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala
    450                 455                 460

CGG GTG AAG AGC GTG CGG GAG GCG GCC GAG CGC ATG GCC TTC AAC ATG   1440
Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met
465                 470                 475                 480

CCC GTC CAG GGC ACC GCC GCC GAC CTC ATG AAG CTG GCT ATG GTG AAG   1488
Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
                485                 490                 495

CTC TTC CCC AGG CTG GAG GAA ATG GGG GCC AGG ATG CTC CTT CAG GTC   1536
Leu Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val
            500                 505                 510

CAC GAC GAG CTG GTC CTC GAG GCC CCA AAA GAG AGG GCG GAG GCC GTG   1584
His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val
        515                 520                 525

GCC CGG CTG GCC AAG GAG GTC ATG GAG GGG GTG TAT CCC CTG GCC GTG   1632
Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val
    530                 535                 540

CCC CTG GAG GTG GAG GTG GGG ATA GGG GAG GAC TGG CTC TCC GCC AAG   1680
Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
545                 550                 555                 560

GAG TGA                                                           1686
Glu  *
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1689 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: FY3
(B) LOCATION: 1...1686

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
ATG GCT CTG GAA CGT CTG GAG TTT GGC AGC CTC CTC CAC GAG TTC GGC   48
```

```
Met Ala Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly
1               5                   10                  15

CTT CTG GAA AGC CCC AAG GCC CTG GAG GAG GCC CCC TGG CCC CCG CCG     96
Leu Leu Glu Ser Pro Lys Ala Leu Glu Glu Ala Pro Trp Pro Pro Pro
            20                  25                  30

GAA GGG GCC TTC GTG GGC TTT GTG CTT TCC CGC AAG GAG CCC ATG TGG    144
Glu Gly Ala Phe Val Gly Phe Val Leu Ser Arg Lys Glu Pro Met Trp
        35                  40                  45

GCC GAT CTT CTG GCC CTG GCC GCC GCC AGG GGG GGC CGG GTC CAC CGG    192
Ala Asp Leu Leu Ala Leu Ala Ala Ala Arg Gly Gly Arg Val His Arg
    50                  55                  60

GCC CCC GAG CCT TAT AAA GCC CTC AGG GAC CTG AAG GAG GCG CGG GGG    240
Ala Pro Glu Pro Tyr Lys Ala Leu Arg Asp Leu Lys Glu Ala Arg Gly
65                  70                  75                  80

CTT CTC GCC AAA GAC CTG AGC GTT CTG GCC CTG AGG GAA GGC CTT GGC    288
Leu Leu Ala Lys Asp Leu Ser Val Leu Ala Leu Arg Glu Gly Leu Gly
                85                  90                  95

CTC CCG CCC GGC GAC GAC CCC ATG CTC CTC GCC TAC CTC CTG GAC CCT    336
Leu Pro Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro
            100                 105                 110

TCC AAC ACC ACC CCC GAG GGG GTG GCC CGG CGC TAC GGC GGG GAG TGG    384
Ser Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp
        115                 120                 125

ACG GAG GAG GCG GGG GAG CGG GCC GCC CTT TCC GAG AGG CTC TTC GCC    432
Thr Glu Glu Ala Gly Glu Arg Ala Ala Leu Ser Glu Arg Leu Phe Ala
    130                 135                 140

AAC CTG TGG GGG AGG CTT GAG GGG GAG GAG AGG CTC CTT TGG CTT TAC    480
Asn Leu Trp Gly Arg Leu Glu Gly Glu Glu Arg Leu Leu Trp Leu Tyr
145                 150                 155                 160

CGG GAG GTG GAG AGG CCC CTT TCC GCT GTC CTG GCC CAC ATG GAG GCC    528
Arg Glu Val Glu Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala
                165                 170                 175

ACG GGG GTG CGC CTG GAC GTG GCC TAT CTC AGG GCC TTG TCC CTG GAG    576
Thr Gly Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu
            180                 185                 190

GTG GCC GAG GAG ATC GCC CGC CTC GAG GCC GAG GTC TTC CGC CTG GCC    624
Val Ala Glu Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala
        195                 200                 205

GGC CAC CCC TTC AAC CTC AAC TCC CGG GAC CAG CTG GAA AGG GTC CTC    672
```

```
Gly His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu
    210                 215                 220

TTT GAC GAG CTA GGG CTT CCC GCC ATC GGC AAG ACG GAG AAG ACC GGC    720
Phe Asp Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly
225                 230                 235                 240

AAG CGC TCC ACC AGC GCC GCC GTC CTG GAG GCC CTC CGC GAG GCC CAC    768
Lys Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His
                245                 250                 255

CCC ATC GTG GAG AAG ATC CTG CAG TAC CGG GAG CTC ACC AAG CTG AAG    816
Pro Ile Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys
            260                 265                 270

AGC ACC TAC ATT GAC CCC TTG CCG GAC CTC ATC CAC CCC AGG ACG GGC    864
Ser Thr Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly
        275                 280                 285

CGC CTC CAC ACC CGC TTC AAC CAG ACG GCC ACG GCC ACG GGC AGG CTA    912
Arg Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu
    290                 295                 300

AGT AGC TCC GAT CCC AAC CTC CAG AAC ATC CCC GTC CGC ACC CCG CTT    960
Ser Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu
305                 310                 315                 320

GGG CAG AGG ATC CGC CGG GCC TTC ATC GCC GAG GAG GGG TGG CTA TTG   1008
Gly Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu
                325                 330                 335

GTG GCC CTG GAC TAT AGC CAG ATA GAG CTC AGG GTG CTG GCC CAC CTC   1056
Val Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu
            340                 345                 350

TCC GGC GAC GAG AAC CTG ATC CGG GTC TTC CAG GAG GGG CGG GAC ATC   1104
Ser Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile
        355                 360                 365

CAC ACG GAG ACC GCC AGC TGG ATG TTC GGC GTC CCC CGG GAG GCC GTG   1152
His Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val
    370                 375                 380

GAC CCC CTG ATG CGC CGG GCG GCC AAG ACC ATC AAC TAC GGG GTC CTC   1200
Asp Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Tyr Gly Val Leu
385                 390                 395                 400

TAC GGC ATG TCG GCC CAC CGC CTC TCC CAG GAG CTA GCC ATC CCT TAC   1248
Tyr Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr
                405                 410                 415
```

```
GAG GAG GCC CAG GCC TTC ATT GAG CGC TAC TTT CAG AGC TTC CCC AAG   1296
Glu Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys
            420                 425                 430

GTG CGG GCC TGG ATT GAG AAG ACC CTG GAG GAG GGC AGG AGG CGG GGG   1344
Val Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly
        435                 440                 445

TAC GTG GAG ACC CTC TTC GGC CGC CGC CGC TAC GTG CCA GAC CTA GAG   1392
Tyr Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu
    450                 455                 460

GCC CGG GTG AAG AGC GTG CGG GAG GCG GCC GAG CGC ATG GCC TTC AAC   1440
Ala Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn
465                 470                 475                 480

ATG CCC GTC CAG GGC ACC GCC GCC GAC CTC ATG AAG CTG GCT ATG GTG   1488
Met Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val
                485                 490                 495

AAG CTC TTC CCC AGG CTG GAG GAA ATG GGG GCC AGG ATG CTC CTT CAG   1536
Lys Leu Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln
            500                 505                 510

GTC CAC GAC GAG CTG GTC CTC GAG GCC CCA AAA GAG AGG GCG GAG GCC   1584
Val His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala
        515                 520                 525

GTG GCC CGG CTG GCC AAG GAG GTC ATG GAG GGG GTG TAT CCC CTG GCC   1632
Val Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala
    530                 535                 540

GTG CCC CTG GAG GTG GAG GTG GGG ATA GGG GAG GAC TGG CTC TCC GCC   1680
Val Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala
545                 550                 555                 560

AAG GAG TGA                                                       1689
Lys Glu  *
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GCTTGGGCAG AGGATCCGCC GGG 23

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 64 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGGATGGCTA GCTCCTGGGA GAGGCGGTGG GCCGACATGC CGTAGAGGAC 50
CCCGTAGTTG ATGG 64

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGAATTCCAT ATGGACGATC TGAAGCTCTC C 31

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GGGGTACCAA GCTTCACTCC TTGGCGGAGA G 31

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GGAATTCCAT ATGCTGGAGA GGCTTGAGTT T 31

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 43 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GGAATTCCAT ATGCTGGAAC GTCTGGAGTT TGGCAGCCTC CTC 43

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GGAATTCCAT ATGGCTCTGG AACGTCTGGA GTTTGGCAGC CTCCTC 46

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GGAATTCCAT ATGCTGGAAC GTCTGGAATT CGGCAGCCTC 40

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GGGGTACCCT AACCCTTGGC GGAAAGCCAG TC 32

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 64 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GGGATGGCTA GCTCCTGGGA GAGCCTATGG GCGGACATGC CGTAGAGGAC 50

GCCGTAGTTC ACCG 64

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

CTAGCTAGCC ATCCCCTACG AAGAAGCGGT GGCCT 35

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1686 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: FY4
(B) LOCATION: 1...1683

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

```
ATG CTG GAA CGT CTG GAA TTC GGC AGC CTC CTC CAC GAG TTC GGC CTC      48
Met Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu
 1               5                  10                  15

CTG GAG GCC CCC GCC CCC CTG GAG GAG GCC CCC TGG CCC CCG CCG GAA      96
Leu Glu Ala Pro Ala Pro Leu Glu Glu Ala Pro Trp Pro Pro Pro Glu
            20                  25                  30

GGG GCC TTC GTG GGC TTC GTC CTC TCC CGC CCC GAG CCC ATG TGG GCG     144
Gly Ala Phe Val Gly Phe Val Leu Ser Arg Pro Glu Pro Met Trp Ala
        35                  40                  45
```

```
GAG CTT AAA GCC CTG GCC GCC TGC AGG GAC GGC CGG GTG CAC CGG GCA     192
Glu Leu Lys Ala Leu Ala Ala Cys Arg Asp Gly Arg Val His Arg Ala
     50                  55                  60

GCA GAC CCC TTG GCG GGG CTA AAG GAC CTC AAG GAG GTC CGG GGC CTC     240
Ala Asp Pro Leu Ala Gly Leu Lys Asp Leu Lys Glu Val Arg Gly Leu
65                  70                  75                  80

CTC GCC AAG GAC CTC GCC GTC TTG GCC TCG AGG GAG GGG CTA GAC CTC     288
Leu Ala Lys Asp Leu Ala Val Leu Ala Ser Arg Glu Gly Leu Asp Leu
                85                  90                  95

GTG CCC GGG GAC GAC CCC ATG CTC CTC GCC TAC CTC CTG GAC CCC TCC     336
Val Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser
            100                 105                 110

AAC ACC ACC CCC GAG GGG GTG GCG CGG CGC TAC GGG GGG GAG TGG ACG     384
Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr
        115                 120                 125

GAG GAC GCC GCC CAC CGG GCC CTC CTC TCG GAG AGG CTC CAT CGG AAC     432
Glu Asp Ala Ala His Arg Ala Leu Leu Ser Glu Arg Leu His Arg Asn
    130                 135                 140

CTC CTT AAG CGC CTC GAG GGG GAG GAG AAG CTC CTT TGG CTC TAC CAC     480
Leu Leu Lys Arg Leu Glu Gly Glu Glu Lys Leu Leu Trp Leu Tyr His
145                 150                 155                 160

GAG GTG GAA AAG CCC CTC TCC CGG GTC CTG GCC CAC ATG GAG GCC ACC     528
Glu Val Glu Lys Pro Leu Ser Arg Val Leu Ala His Met Glu Ala Thr
                165                 170                 175

GGG GTA CGG CTG GAC GTG GCC TAC CTT CAG GCC CTT TCC CTG GAG CTT     576
Gly Val Arg Leu Asp Val Ala Tyr Leu Gln Ala Leu Ser Leu Glu Leu
            180                 185                 190

GCG GAG GAG ATC CGC CGC CTC GAG GAG GAG GTC TTC CGC TTG GCG GGC     624
Ala Glu Glu Ile Arg Arg Leu Glu Glu Glu Val Phe Arg Leu Ala Gly
        195                 200                 205

CAC CCC TTC AAC CTC AAC TCC CGG GAC CAG CTG GAA AGG GTG CTC TTT     672
His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe
    210                 215                 220

GAC GAG CTT AGG CTT CCC GCC TTG GGG AAG ACG CAA AAG ACA GGC AAG     720
Asp Glu Leu Arg Leu Pro Ala Leu Gly Lys Thr Gln Lys Thr Gly Lys
225                 230                 235                 240

CGC TCC ACC AGC GCC GCG GTG CTG GAG GCC CTA CGG GAG GCC CAC CCC     768
Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro
                245                 250                 255

ATC GTG GAG AAG ATC CTC CAG CAC CGG GAG CTC ACC AAG CTC AAG AAC     816
```

```
Ile Val Glu Lys Ile Leu Gln His Arg Glu Leu Thr Lys Leu Lys Asn
            260                     265                     270

ACC TAC GTG GAC CCC CTC CCA AGC CTC GTC CAC CCG AGG ACG GGC CGC      864
Thr Tyr Val Asp Pro Leu Pro Ser Leu Val His Pro Arg Thr Gly Arg
        275                     280                     285

CTC CAC ACC CGC TTC AAC CAG ACG GCC ACG GCC ACG GGG AGG CTT AGT      912
Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser
    290                     295                     300

AGC TCC GAC CCC AAC CTG CAG AAC ATC CCC GTC CGC ACC CCC TTG GGC      960
Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly
305                     310                     315                 320

CAG AGG ATC CGC CGG GCC TTC GTG GCC GAG GCG GGT TGG GCG TTG GTG     1008
Gln Arg Ile Arg Arg Ala Phe Val Ala Glu Ala Gly Trp Ala Leu Val
                325                     330                     335

GCC CTG GAC TAT AGC CAG ATA GAG CTC CGC GTC CTC GCC CAC CTC TCC     1056
Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser
            340                     345                     350

GGG GAC GAA AAC CTG ATC AGG GTC TTC CAG GAG GGG AAG GAC ATC CAC     1104
Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Lys Asp Ile His
        355                     360                     365

ACC CAG ACC GCA AGC TGG ATG TTC GGC GTC CCC CCG GAG GCC GTG GAC     1152
Thr Gln Thr Ala Ser Trp Met Phe Gly Val Pro Pro Glu Ala Val Asp
    370                     375                     380

CCC CTG ATG CGC CGG GCG GCC AAG ACG GTG AAC TAC GGC GTC CTC TAC     1200
Pro Leu Met Arg Arg Ala Ala Lys Thr Val Asn Tyr Gly Val Leu Tyr
385                     390                     395                 400

GGC ATG TCC GCC CAT AGG CTC TCC CAG GAG CTA GCC ATC CCC TAC GAA     1248
Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu
                405                     410                     415

GAA GCG GTG GCC TTT ATA GAG CGC TAC TTC CAA AGC TTC CCC AAG GTG     1296
Glu Ala Val Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val
            420                     425                     430

CGG GCC TGG ATA GAA AAG ACC CTG GAG GAG GGG AGG AAG CGG GGC TAC     1344
Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Lys Arg Gly Tyr
        435                     440                     445

GTG GAA ACC CTC TTC GGA AGA AGG CGC TAC GTG CCC GAC CTC AAC GCC     1392
Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Asn Ala
    450                     455                     460

CGG GTG AAG AGC GTC AGG GAG GCC GCG GAG CGC ATG GCC TTC AAC ATG     1440
Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met
```

```
465                 470                 475                 480

CCC GTC CAG GGC ACC GCC GCC GAC CTC ATG AAG CTC GCC ATG GTG AAG    1488
Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
                485                 490                 495

CTC TTC CCC CGC CTC CGG GAG ATG GGG GCC CGC ATG CTC CTC CAG GTC    1536
Leu Phe Pro Arg Leu Arg Glu Met Gly Ala Arg Met Leu Leu Gln Val
            500                 505                 510

CAC GAC GAG CTC CTC CTG GAG GCC CCC CAA GCG CGG GCC GAG GAG GTG    1584
His Asp Glu Leu Leu Leu Glu Ala Pro Gln Ala Arg Ala Glu Glu Val
        515                 520                 525

GCG GCT TTG GCC AAG GAG GCC ATG GAG AAG GCC TAT CCC CTC GCC GTG    1632
Ala Ala Leu Ala Lys Glu Ala Met Glu Lys Ala Tyr Pro Leu Ala Val
    530                 535                 540

CCC CTG GAG GTG GAG GTG GGG ATG GGG GAG GAC TGG CTT TCC GCC AAG    1680
Pro Leu Glu Val Glu Val Gly Met Gly Glu Asp Trp Leu Ser Ala Lys
545                 550                 555                 560

GGT TAG                                                             1686
Gly  *
```

Claims

1. An enzymatically active DNA polymerase having between 540 and 582 amino acids having a tyrosine at a position equivalent to position 667 of Taq DNA polymerase, wherein said polymerase lacks 5' to 3' exonuclease activity, and wherein said polymerase has at least 95% homology in its amino acid sequence to the DNA polymerase of Thermus aquaticus, Thermus flavus or Thermus thermophilus, and wherein said polymerase forms a single polypeptide band or an SDS polyacrylamide gel.

2. The polymerase of claim 1 wherein the amino acid sequence of said polymerase includes less than 3 conservative amino acid changes compared to one said DNA polymerase of said named Thermus species.

3. The polymerase of claim 1 wherein the amino acid sequence of said polymerase includes less than 3 additional amino acids compared to one said DNA polymerase of said named Thermus species at its N-terminus.

4. The polymerase of claim 1 selected from the group consisting of FY2, FY3 and FY4.

5. Purified nucleic acid encoding the DNA polymerase of any of claims 1-4.

6. Method for sequencing DNA comprising the step of generating chain terminated fragments from the DNA template to be sequenced with a DNA polymerase of any of claims 1-4 in the presence of at least one chain terminating agent and one or more nucleotide triphosphates, and determining the sequence of said DNA from the sizes of said fragments.

7. Kit for sequencing DNA comprising a DNA polymerase of any of claims 1-4 and a pyrophosphatase.

8. The kit of claim 7 wherein said pyrophosphatase is thermostable.

9. Apparatus for DNA sequencing having a reactor comprising a DNA polymerase of any of claims 1-4 and a band separator.

FIG. 1
(sheet 1)

```
DNA  sequence   1686  b.p.   atgctggagagg ... gccaaggagtga   linear

1/1                                     31/11
atg ctg gag agg ctt gag ttt ggc agc ctc ctc cac gag ttc ggc ctt ctg gaa agc ccc
M   L   E   R   L   E   F   G   S   L   L   H   E   F   G   L   L   E   S   P
61/21                                   91/31
aag gcc ctg gag gag gcc ccc tgg ccc ccg ccg gaa ggg gcc ttc gtg ggc ttt gtg ctt
K   A   L   E   E   A   P   W   P   P   P   E   G   A   F   V   G   F   V   L
121/41                                  151/51
tcc cgc aag gag ccc atg tgg gcc gat ctt ctg gcc ctg gcc gcc gcc agg ggg ggc cgg
S   R   K   E   P   M   W   A   D   L   L   A   L   A   A   A   R   G   G   R
181/61                                  211/71
gtc cac cgg gcc ccc gag cct tat aaa gcc ctc agg gac ctg aag gag gcg cgg ggg ctt
V   H   R   A   P   E   P   Y   K   A   L   R   D   L   K   E   A   R   G   L
241/81                                  271/91
ctc gcc aaa gac ctg agc gtt ctg gcc ctg agg gaa ggc ctt ggc ctc ccg ccc ggc gac
L   A   K   D   L   S   V   L   A   L   R   E   G   L   G   L   P   P   G   D
301/101                                 331/111
gac ccc atg ctc ctc gcc tac ctc ctg gac cct tcc aac acc acc ccc gag ggg gtg gcc
D   P   M   L   L   A   Y   L   L   D   P   S   N   T   T   P   E   G   V   A
361/121                                 391/131
cgg cgc tac ggc ggg gag tgg acg gag gag gcg ggg gag cgg gcc gcc ctt tcc gag agg
R   R   Y   G   G   E   W   T   E   E   A   G   E   R   A   A   L   S   E   R
421/141                                 451/151
ctc ttc gcc aac ctg tgg ggg agg ctt gag ggg gag gag agg ctc ctt tgg ctt tac cgg
L   F   A   N   L   W   G   R   L   E   G   E   E   R   L   L   W   L   Y   R
481/161                                 511/171
gag gtg gag agg ccc ctt tcc gct gtc ctg gcc cac atg gag gcc acg ggg gtg cgc ctg
E   V   E   R   P   L   S   A   V   L   A   H   M   E   A   T   G   V   R   L
541/181                                 571/191
gac gtg gcc tat ctc agg gcc ttg tcc ctg gag gtg gcc gag gag atc gcc cgc ctc gag
D   V   A   Y   L   R   A   L   S   L   E   V   A   E   E   I   A   R   L   E
601/201                                 631/211
gcc gag gtc ttc cgc ctg gcc ggc cac ccc ttc aac ctc aac tcc cgg gac cag ctg gaa
A   E   V   F   R   L   A   G   H   P   F   N   L   N   S   R   D   Q   L   E
661/221                                 691/231
agg gtc ctc ttt gac gag cta ggg ctt ccc gcc atc ggc aag acg gag aag acc ggc aag
R   V   L   F   D   E   L   G   L   P   A   I   G   K   T   E   K   T   G   K
721/241                                 751/251
cgc tcc acc agc gcc gcc gtc ctg gag gcc ctc cgc gag gcc cac ccc atc gtg gag aag
R   S   T   S   A   A   V   L   E   A   L   R   E   A   H   P   I   V   E   K
781/261                                 811/271
atc ctg cag tac cgg gag ctc acc aag ctg aag agc acc tac att gac ccc ttg ccg gac
I   L   Q   Y   R   E   L   T   K   L   K   S   T   Y   I   D   P   L   P   D
841/281                                 871/291
ctc atc cac ccc agg acg ggc cgc ctc cac acc cgc ttc aac cag acg gcc acg gcc acg
L   I   H   P   R   T   G   R   L   H   T   R   F   N   Q   T   A   T   A   T
901/301                                 931/311
ggc agg cta agt agc tcc gat ccc aac ctc cag aac atc ccc gtc cgc acc ccg ctt ggg
G   R   L   S   S   S   D   P   N   L   Q   N   I   P   V   R   T   P   L   G
961/321                                 991/331
cag agg atc cgc cgg gcc ttc atc gcc gag gag ggg tgg cta ttg gtg gcc ctg gac tat
Q   R   I   R   R   A   F   I   A   E   E   G   W   L   L   V   A   L   D   Y
1021/341                                1051/351
agc cag ata gag ctc agg gtg ctg gcc cac ctc tcc ggc gac gag aac ctg atc cgg gtc
S   Q   I   E   L   R   V   L   A   H   L   S   G   D   E   N   L   I   R   V
1081/361                                1111/371
ttc cag gag ggg cgg gac atc cac acg gag acc gcc agc tgg atg ttc ggc gtc ccc cgg
F   Q   E   G   R   D   I   H   T   E   T   A   S   W   M   F   G   V   P   R
1141/381                                1171/391
gag gcc gtg gac ccc ctg atg cgc cgg gcg gcc aag acc atc aac tac ggg gtc ctc tac
E   A   V   D   P   L   M   R   R   A   A   K   T   I   N   Y   G   V   L   Y
1201/401                                1231/411
ggc atg tcg gcc cac cgc ctc tcc cag gag cta gcc atc cct tac gag gag gcc cag gcc
G   M   S   A   H   R   L   S   Q   E
```

```
1261/421                                1291/431
ttc att gag cgc tac ttt cag agc ttc ccc aag gtg cgg gcc tgg att gag aag acc ctg
F   I   E   R   Y   F   Q   S   F   P   K   V   R   A   W   I   E   K   T   L
1321/441                                1351/451
gag gag ggc agg agg cgg ggg tac gtg gag acc ctc ttc ggc cgc cgc cgc tac gtg cca
E   E   G   R   R   R   G   Y   V   E   T   L   F   G   R   R   R   Y   V   P
1381/461                                1411/471
gac cta gag gcc cgg gtg aag agc gtg cgg gag gcg gcc gag cgc atg gcc ttc aac atg
D   L   E   A   R   V   K   S   V   R   E   A   A   E   R   M   A   F   N   M
1441/481                                1471/491
ccc gtc cag ggc acc gcc gcc gac ctc atg aag ctg gct atg gtg aag ctc ttc ccc agg
P   V   Q   G   T   A   A   D   L   M   K   L   A   M   V   K   L   F   P   R
1501/501                                1531/511
ctg gag gaa atg ggg gcc agg atg ctc ctt cag gtc cac gac gag ctg gtc ctc gag gcc
L   E   E   M   G   A   R   M   L   L   Q   V   H   D   E   L   V   L   E   A
1561/521                                1591/531
cca aaa gag agg gcg gag gcc gtg gcc cgg ctg gcc aag gag gtc atg gag ggg gtg tat
P   K   E   R   A   E   A   V   A   R   L   A   K   E   V   M   E   G   V   Y
1621/541                                1651/551
ccc ctg gcc gtg ccc ctg gag gtg gag gtg ggg ata ggg gag gac tgg ctc tcc gcc aag
P   L   A   V   P   L   E   V   E   V   G   I   G   E   D   W   L   S   A   K
1681/561
gag tga
E   *
```

FIG. 1
(sheet 2)

FIG. 2
(sheet 1)

```
DNA  sequence    1689  b.p.   atggctctggaa ... gccaaggagtga   linear

1/1                                     31/11
atg gct ctg gaa cgt ctg gag ttt ggc agc ctc ctc cac gag ttc ggc ctt ctg gaa agc
M   A   L   E   R   L   E   F   G   S   L   L   H   E   F   G   L   L   E   S
61/21                                   91/31
ccc aag gcc ctg gag gag gcc ccc tgg ccc ccg ccg gaa ggg gcc ttc gtg ggc ttt gtg
P   K   A   L   E   E   A   P   W   P   P   P   E   G   A   F   V   G   F   V
121/41                                  151/51
ctt tcc cgc aag gag ccc atg tgg gcc gat ctt ctg gcc ctg gcc gcc gcc agg ggg ggc
L   S   R   K   E   P   M   W   A   D   L   L   A   L   A   A   A   R   G   G
181/61                                  211/71
cgg gtc cac cgg gcc ccc gag cct tat aaa gcc ctc agg gac ctg aag gag gcg cgg ggg
R   V   H   R   A   P   E   P   Y   K   A   L   R   D   L   K   E   A   R   G
241/81                                  271/91
ctt ctc gcc aaa gac ctg agc gtt ctg gcc ctg agg gaa ggc ctt ggc ctc ccg ccc ggc
L   L   A   K   D   L   S   V   L   A   L   R   E   G   L   G   L   P   P   G
301/101                                 331/111
gac gac ccc atg ctc ctc gcc tac ctc ctg gac cct tcc aac acc acc ccc gag ggg gtg
D   D   P   M   L   L   A   Y   L   L   D   P   S   N   T   T   P   E   G   V
361/121                                 391/131
gcc cgg cgc tac ggc ggg gag tgg acg gag gag gcg ggg gag cgg gcc gcc ctt tcc gag
A   R   R   Y   G   G   E   W   T   E   E   A   G   E   R   A   A   L   S   E
421/141                                 451/151
agg ctc ttc gcc aac ctg tgg ggg agg ctt gag ggg gag gag agg ctc ctt tgg ctt tac
R   L   F   A   N   L   W   G   R   L   E   G   E   E   R   L   L   W   L   Y
481/161                                 511/171
cgg gag gtg gag agg ccc ctt tcc gct gtc ctg gcc cac atg gag gcc acg ggg gtg cgc
R   E   V   E   R   P   L   S   A   V   L   A   H   M   E   A   T   G   V   R
541/181                                 571/191
ctg gac gtg gcc tat ctc agg gcc ttg tcc ctg gag gtg gcc gag gag atc gcc cgc ctc
L   D   V   A   Y   L   R   A   L   S   L   E   V   A   E   E   I   A   R   L
601/201                                 631/211
gag gcc gag gtc ttc cgc ctg gcc ggc cac ccc ttc aac ctc aac tcc cgg gac cag ctg
E   A   E   V   F   R   L   A   G   H   P   F   N   L   N   S   R   D   Q   L
661/221                                 691/231
gaa agg gtc ctc ttt gac gag cta ggg ctt ccc gcc atc ggc aag acg gag aag acc ggc
E   R   V   L   F   D   E   L   G   L   P   A   I   G   K   T   E   K   T   G
721/241                                 751/251
aag cgc tcc acc agc gcc gcc gtc ctg gag gcc ctc cgc gag gcc cac ccc atc gtg gag
K   R   S   T   S   A   A   V   L   E   A   L   R   E   A   H   P   I   V   E
781/261                                 811/271
aag atc ctg cag tac cgg gag ctc acc aag ctg aag agc acc tac att gac ccc ttg ccg
K   I   L   Q   Y   R   E   L   T   K   L   K   S   T   Y   I   D   P   L   P
841/281                                 871/291
gac ctc atc cac ccc agg acg ggc cgc ctc cac acc cgc ttc aac cag acg gcc acg gcc
D   L   I   H   P   R   T   G   R   L   H   T   R   F   N   Q   T   A   T   A
901/301                                 931/311
acg ggc agg cta agt agc tcc gat ccc aac ctc cag aac atc ccc gtc cgc acc ccg ctt
T   G   R   L   S   S   S   D   P   N   L   Q   N   I   P   V   R   T   P   L
961/321                                 991/331
ggg cag agg atc cgc cgg gcc ttc atc gcc gag gag ggg tgg cta ttg gtg gcc ctg gac
G   Q   R   I   R   R   A   F   I   A   E   E   G   W   L   L   V   A   L   D
1021/341                                1051/351
tat agc cag ata gag ctc agg gtg ctg gcc cac ctc tcc ggc gac gag aac ctg atc cgg
Y   S   Q   I   E   L   R   V   L   A   H   L   S   G   D   E   N   L   I   R
1081/361                                1111/371
gtc ttc cag gag ggg cgg gac atc cac acg gag acc gcc agc tgg atg ttc ggc gtc ccc
V   F   Q   E   G   R   D   I   H   T   E   T   A   S   W   M   F   G   V   P
1141/381                                1171/391
cgg gag gcc gtg gac ccc ctg atg cgc cgg gcg gcc aag acc atc aac tac ggg gtc ctc
R   E   A   V   D   P   L   M   R   R   A   A   K   T   I   N   Y   G   V   L
1201/401                                1231/411
tac ggc atg tcg gcc cac cgc ctc tcc cag gag cta gcc atc cct tac gag gag gcc cag
Y   G   M   S   A   H   R   L   S   Q   E   L   A   I   P   Y   E   E   A   Q
```

```
1261/421                                          1291/431
gcc ttc att gag cgc tac ttt cag agc ttc ccc aag gtg cgg gcc tgg att gag aag acc
A   F   I   E   R   Y   F   Q   S   F   P   K   V   R   A   W   I   E   K   T
1321/441                                          1351/451
atg gag gag ggc agg agg cgg ggg tac gtg gag acc ctc ttc ggc cgc cgc cgc tac gtg
M   E   E   G   R   R   R   G   Y   V   E   T   L   F   G   R   R   R   Y   V
1381/461                                          1411/471
cca gac cta gag gcc cgg gtg aag agc gtg cgg gag gcg gcc gag cgc atg gcc ttc aac
P   D   L   E   A   R   V   K   S   V   R   E   A   A   E   R   M   A   F   N
1441/481                                          1471/491
atg ccc gtc cag ggc acc gcc gcc gac ctc atg aag ctg gct atg gtg aag ctc ttc ccc
M   P   V   Q   G   T   A   A   D   L   M   K   L   A   M   V   K   L   F   P
1501/501                                          1531/511
agg ctg gag gaa atg ggg gcc agg atg ctc ctt cag gtc cac gac gag ctg gtc ctc gag
R   L   E   E   M   G   A   R   M   L   L   Q   V   H   D   E   L   V   L   E
1561/521                                          1591/531
gcc cca aaa gag agg gcg gag gcc gtg gcc cgg ctg gcc aag gag gtc atg gag ggg gtg
A   P   K   E   R   A   E   A   V   A   R   L   A   K   E   V   M   E   G   V
1621/541                                          1651/551
tat ccc ctg gcc gtg ccc ctg gag gtg gag gtg ggg ata ggg gag gac tgg ctc tcc gcc
Y   P   L   A   V   P   L   E   V   E   V   G   I   G   E   D   W   L   S   A
1681/561
aag gag tga
K   E   *
```

FIG. 2
(sheet 2)

FIG. 3
(sheet 1)

```
DNA  sequence     2496  b.p.   atggcgatgctt ... gccaaggagtag   linear

1/1                                     31/11
atg gcg atg ctt ccc ctc ttt gag ccc aaa ggc cgc gtg ctc ctg gtg gac ggc cac cac
M   A   M   L   P   L   F   E   P   K   G   R   V   L   L   V   D   G   H   H
61/21                                   91/31
ctg gcc tac cgc acc ttc ttt gcc ctc aag ggc ctc acc acc agc cgc ggc gaa ccc gtt
L   A   Y   R   T   F   F   A   L   K   G   L   T   T   S   R   G   E   P   V
121/41                                  151/51
cag gcg gtc tac ggc ttc gcc aaa agc ctc ctc aag gcc ctg aag gag gac ggg gac gtg
Q   A   V   Y   G   F   A   K   S   L   L   K   A   L   K   E   D   G   D   V
181/61                                  211/71
gtg gtg gtg gtc ttt gac gcc aag gcc ccc tcc ttc cgc cac gag gcc tac gag gcc tac
V   V   V   V   F   D   A   K   A   P   S   F   R   H   E   A   Y   E   A   Y
241/81                                  271/91
aag gcg ggc cgg gcc ccc acc ccg gag gac ttt ccc cgg cag ctg gcc ctc atc aag gag
K   A   G   R   A   P   T   P   E   D   F   P   R   Q   L   A   L   I   K   E
301/101                                 331/111
ttg gtg gac ctc cta ggc ctt gtg cgg ctg gag gtt ccc ggc ttt gag gcg gac gac gtg
L   V   D   L   L   G   L   V   R   L   E   V   P   G   F   E   A   D   D   V
361/121                                 391/131
ctg gcc acc ctg gcc aag cgg gcg gaa aag gag ggg tac gag gtg cgc atc ctc act gcc
L   A   T   L   A   K   R   A   E   K   E   G   Y   E   V   R   I   L   T   A
421/141                                 451/151
gac cgc gac ctc tac cag ctc ctt tcg gag cgc atc gcc atc ctc cac cct gag ggg tac
D   R   D   L   Y   Q   L   L   S   E   R   I   A   I   L   H   P   E   G   Y
481/161                                 511/171
ctg atc acc ccg gcg tgg ctt tac gag aag tac ggc ctg cgc ccg gag cag tgg gtg gac
L   I   T   P   A   W   L   Y   E   K   Y   G   L   R   P   E   Q   W   V   D
541/181                                 571/191
tac cgg gcc ctg gcg ggg gac ccc tcg gat aac atc ccc ggg gtg aag ggc atc ggg gag
Y   R   A   L   A   G   D   P   S   D   N   I   P   G   V   K   G   I   G   E
601/201                                 631/211
aag acc gcc cag agg ctc atc cgc gag tgg ggg agc ctg gaa aac ctc ttc cag cac ctg
K   T   A   Q   R   L   I   R   E   W   G   S   L   E   N   L   F   Q   H   L
661/221                                 691/231
gac cag gtg aag ccc tcc ttg cgg gag aag ctc cag gcg ggc atg gag gcc ctg gcc ctt
D   Q   V   K   P   S   L   R   E   K   L   Q   A   G   M   E   A   L   A   L
721/241                                 751/251
tcc cgg aag ctt tcc cag gtg cac act gac ctg ccc ctg gag gtg gac ttc ggg agg cgc
S   R   K   L   S   Q   V   H   T   D   L   P   L   E   V   D   F   G   R   R
781/261                                 811/271
cgc aca ccc aac ctg gag ggt ctg cgg gct ttt ttg gag cgg ttg gag ttt gga agc ctc
R   T   P   N   L   E   G   L   R   A   F   L   E   R   L   E   F   G   S   L
841/281                                 871/291
ctc cac gag ttc ggc ctc ctg gag ggg ccg aag gcg gca gag gag gcc ccc tgg ccc cct
L   H   E   F   G   L   L   E   G   P   K   A   A   E   E   A   P   W   P   P
901/301                                 931/311
ccg gaa ggg gct ttt ttg ggc ttt tcc ttt tcc cgt ccc gag ccc atg tgg gcc gag ctt
P   E   G   A   F   L   G   F   S   F   S   R   P   E   P   M   W   A   E   L
961/321                                 991/331
ctg gcc ctg gct ggg gcg tgg gag ggg cgc ctc cat cgg gca caa gac ccc ctt agg ggc
L   A   L   A   G   A   W   E   G   R   L   H   R   A   Q   D   P   L   R   G
1021/341                                1051/351
ctg agg gac ctt aag ggg gtg cgg gga atc ctg gcc aag gac ctg gcg gtt ttg gcc ctg
L   R   D   L   K   G   V   R   G   I   L   A   K   D   L   A   V   L   A   L
1081/361                                1111/371
cgg gag ggc ctg gac ctc ttc cca gag gac gac ccc atg ctc ctg gcc tac ctt ctg gac
R   E   G   L   D   L   F   P   E   D   D   P   M   L   L   A   Y   L   L   D
1141/381                                1171/391
ccc tcc aac acc acc cct gag ggg gtg gcc cgg cgt tac ggg ggg gag tgg acg gag gat
P   S   N   T   T   P   E   G   V   A   R   R   Y   G   G   E   W   T   E   D
1201/401                                1231/411
gcg ggg gag agg gcc ctc ctg gcc gag cgc ctc ttc cag acc cta aag gag cgc ctt aag
A   G   E   R   A   L   L   A   E   R   L   F   Q   T   L   K   E   R   L   K
```

```
1261/421                                1291/431
gga gaa gaa cgc ctg ctt tgg ctt tac gag gag gtg gag aag ccg ctt tcc cgg gtg ttg
G   E   E   R   L   L   W   L   Y   E   E   V   E   K   P   L   S   R   V   L
1321/441                                1351/451
gcc cgg atg gag gcc acg ggg gtc cgg ctg gac gtg gcc tac ctc cag gcc ctc tcc ctg
A   R   M   E   A   T   G   V   R   L   D   V   A   Y   L   Q   A   L   S   L
1381/461                                1411/471
gag gtg gag gcg gag gtg cgc cag ctg gag gag gag gtc ttc cgc ctg gcc ggc cac ccc
E   V   E   A   E   V   R   Q   L   E   E   E   V   F   R   L   A   G   H   P
1441/481                                1471/491
ttc aac ctc aac tcc cgc gac cag ctg gag cgg gtg ctc ttt gac gag ctg ggc ctg cct
F   N   L   N   S   R   D   Q   L   E   R   V   L   F   D   E   L   G   L   P
1501/501                                1531/511
gcc atc ggc aag acg gag aag acg ggg aaa cgc tcc acc agc gct gcc gtg ctg gag gcc
A   I   G   K   T   E   K   T   G   K   R   S   T   S   A   A   V   L   E   A
1561/521                                1591/531
ctg cga gag gcc cac ccc atc gtg gac cgc atc ctg cag tac cgg gag ctc acc aag ctc
L   R   E   A   H   P   I   V   D   R   I   L   Q   Y   R   E   L   T   K   L
1621/541                                1651/551
aag aac acc tac ata gac ccc ctg ccc gcc ctg gtc cac ccc aag acc ggc cgg ctc cac
K   N   T   Y   I   D   P   L   P   A   L   V   H   P   K   T   G   R   L   H
1681/561                                1711/571
acc cgc ttc aac cag acg gcc acc gcc acg ggc agg ctt tcc agc tcc gac ccc aac ctg
T   R   F   N   Q   T   A   T   A   T   G   R   L   S   S   S   D   P   N   L
1741/581                                1771/591
cag aac atc ccc gtg cgc acc cct ctg ggc cag cgc atc cgc cga gcc ttc gtg gcc gag
Q   N   I   P   V   R   T   P   L   G   Q   R   I   R   R   A   F   V   A   E
1801/601                                1831/611
gag ggc tgg gtg ctg gtg gtc ttg gac tac agc cag att gag ctt cgg gtc ctg gcc cac
E   G   W   V   L   V   V   L   D   Y   S   Q   I   E   L   R   V   L   A   H
1861/621                                1891/631
ctc tcc ggg gac gag aac ctg atc cgg gtc ttt cag gag ggg agg gac atc cac acc cag
L   S   G   D   E   N   L   I   R   V   F   Q   E   G   R   D   I   H   T   Q
1921/641                                1951/651
acc gcc agc tgg atg ttc ggc gtt tcc ccc gaa ggg gta gac cct ctg atg cgc cgg gcg
T   A   S   W   M   F   G   V   S   P   E   G   V   D   P   L   M   R   R   A
1981/661                                2011/671
gcc aag acc atc aac ttc ggg gtg ctc tac ggc atg tcc gcc cac cgc ctc tcc ggg gag
A   K   T   I   N   F   G   V   L   Y   G   M   S   A   H   R   L   S   G   E
2041/681                                2071/691
ctt tcc atc ccc tac gag gag gcg gtg gcc ttc att gag cgc tac ttc cag agc tac ccc
L   S   I   P   Y   E   E   A   V   A   F   I   E   R   Y   F   Q   S   Y   P
2101/701                                2131/711
aag gtg cgg gcc tgg att gag ggg acc ctc gag gag ggc cgc cgg cgg ggg tat gtg gag
K   V   R   A   W   I   E   G   T   L   E   E   G   R   R   R   G   Y   V   E
2161/721                                2191/731
acc ctc ttc ggc cgc cgg cgc tat gtg ccc gac ctc aac gcc cgg gtg aag agc gtg cgc
T   L   F   G   R   R   R   Y   V   P   D   L   N   A   R   V   K   S   V   R
2221/741                                2251/751
gag gcg gcg gag cgc atg gcc ttc aac atg ccg gtc cag ggc acc gcc gcc gac ctc atg
E   A   A   E   R   M   A   F   N   M   P   V   Q   G   T   A   A   D   L   M
2281/761                                2311/771
aag ctg gcc atg gtg cgg ctt ttc ccc cgg ctt cag gaa ctg ggg gcg agg atg ctt ttg
K   L   A   M   V   R   L   F   P   R   L   Q   E   L   G   A   R   M   L   L
2341/781                                2371/791
cag gtg cac gac gag ctg gtc ctc gag gcc ccc aag gac cgg gcg gag agg gta gcc gct
Q   V   H   D   E   L   V   L   E   A   P   K   D   R   A   E   R   V   A   A
2401/801                                2431/811
ttg gcc aag gag gtc atg gag ggg gtc tgg ccc ctg cag gtg ccc ctg gag gtg gag gtg
L   A   K   E   V   M   E   G   V   W   P   L   Q   V   P   L   E   V   E   V
2461/821                                2491/831
ggc ctg ggg gag gac tgg ctc tcc gcc aag gag tag
G   L   G   E   D   W   L   S   A   K   E   •
```

FIG. 3
(sheet 2)

FIG. 4
(sheet 1)

DNA sequence 2505 b.p. ATGGAGGCGATG ... GCCAAGGGTTAG linear

coding sequence of T. thermophilus DNA polymerase as submitted by D. Gelfand in WO 91/09950 PCT/US90/07

```
1/1                                     31/11
ATG GAG GCG ATG CTT CCG CTC TTT GAA CCC AAA GGC CGG GTC CTC CTG GTG GAC GGC CAC
M   E   A   M   L   P   L   F   E   P   K   G   R   V   L   L   V   D   G   H
61/21                                   91/31
CAC CTG GCC TAC CGC ACC TTC TTC GCC CTG AAG GGC CTC ACC ACG AGC CGG GGC GAA CCG
H   L   A   Y   R   T   F   F   A   L   K   G   L   T   T   S   R   G   E   P
121/41                                  151/51
GTG CAG GCG GTC TAC GGC TTC GCC AAG AGC CTC CTC AAG GCC CTG AAG GAG GAC GGG TAC
V   Q   A   V   Y   G   F   A   K   S   L   L   K   A   L   K   E   D   G   Y
181/61                                  211/71
AAG GCC GTC TTC GTG GTC TTT GAC GCC AAG GCC CCC TCC TTC CGC CAC GAG GCC TAC GAG
K   A   V   F   V   V   F   D   A   K   A   P   S   F   R   H   E   A   Y   E
241/81                                  271/91
GCC TAC AAG GCG GGG AGG GCC CCG ACC CCC GAG GAC TTC CCC CGG CAG CTC GCC CTC ATC
A   Y   K   A   G   R   A   P   T   P   E   D   F   P   R   Q   L   A   L   I
301/101                                 331/111
AAG GAG CTG GTG GAC CTC CTG GGG TTT ACC CGC CTC GAG GTC CCC GGC TAC GAG GCG GAC
K   E   L   V   D   L   L   G   F   T   R   L   E   V   P   G   Y   E   A   D
361/121                                 391/131
GAC GTT CTC GCC ACC CTG GCC AAG AAG GCG GAA AAG GAG GGG TAC GAG GTG CGC ATC CTC
D   V   L   A   T   L   A   K   K   A   E   K   E   G   Y   E   V   R   I   L
421/141                                 451/151
ACC GCC GAC CGC GAC CTC TAC CAA CTC GTC TCC GAC CGC GTC GCC GTC CTC CAC CCC GAG
T   A   D   R   D   L   Y   Q   L   V   S   D   R   V   A   V   L   H   P   E
481/161                                 511/171
GGC CAC CTC ATC ACC CCG GAG TGG CTT TGG GAG AAG TAC GGC CTC AGG CCG GAG CAG TGG
G   H   L   I   T   P   E   W   L   W   E   K   Y   G   L   R   P   E   Q   W
541/181                                 571/191
GTG GAC TTC CGC GCC CTC GTG GGG GAC CCC TCC GAC AAC CTC CCC GGG GTC AAG GGC ATC
V   D   F   R   A   L   V   G   D   P   S   D   N   L   P   G   V   K   G   I
601/201                                 631/211
GGG GAG AAG ACC GCC CTC AAG CTC CTC AAG GAG TGG GGA AGC CTG GAA AAC CTC CTC AAG
G   E   K   T   A   L   K   L   L   K   E   W   G   S   L   E   N   L   L   K
661/221                                 691/231
AAC CTG GAC CGG GTA AAG CCA GAA AAC GTC CGG GAG AAG ATC AAG GCC CAC CTG GAA GAC
N   L   D   R   V   K   P   E   N   V   R   E   K   I   K   A   H   L   E   D
721/241                                 751/251
CTC AGG CTC TCC TTG GAG CTC TCC CGG GTG CGC ACC GAC CTC CCC CTG GAG GTG GAC CTC
L   R   L   S   L   E   L   S   R   V   R   T   D   L   P   L   E   V   D   L
781/261                                 811/271
GCC CAG GGG CGG GAG CCC GAC CGG GAG GGG CTT AGG GCC TTC CTG GAG AGG CTG GAG TTC
A   Q   G   R   E   P   D   R   E   G   L   R   A   F   L   E   R   L   E   F
841/281                                 871/291
GGC AGC CTC CTC CAC GAG TTC GGC CTC CTG GAG GCC CCC GCC CCC CTG GAG GAG GCC CCC
G   S   L   L   H   E   F   G   L   L   E   A   P   A   P   L   E   E   A   P
901/301                                 931/311
TGG CCC CCG CCG GAA GGG GCC TTC GTG GGC TTC GTC CTC TCC CGC CCC GAG CCC ATG TGG
W   P   P   P   E   G   A   F   V   G   F   V   L   S   R   P   E   P   M   W
961/321                                 991/331
GCG GAG CTT AAA GCC CTG GCC GCC TGC AGG GAC GGC CGG GTG CAC CGG GCA GCA GAC CCC
A   E   L   K   A   L   A   A   C   R   D   G   R   V   H   R   A   A   D   P
1021/341                                1051/351
TTG GCG GGG CTA AAG GAC CTC AAG GAG GTC CGG GGC CTC CTC GCC AAG GAC CTC GCC GTC
L   A   G   L   K   D   L   K   E   V   R   G   L   L   A   K   D   L   A   V
1081/361                                1111/371
TTG GCC TCG AGG GAG GGG CTA GAC CTC GTG CCC GGG GAC GAC CCC ATG CTC CTC GCC TAC
L   A   S   R   E   G   L   D   L   V   P   G   D   D   P   M   L   L   A   Y
1141/381                                1171/391
CTC CTG GAC CCC TCC AAC ACC ACC CCC GAG GGG GTG GCG CGG CGC TAC GGG GGG GAG TGG
L   L   D   P   S   N   T   T   P   E   G   V   A   R   R   Y   G   G   E   W
```

1201/401 1231/411
ACG GAG GAC GCC GCC CAC CGG GCC CTC CTC TCG GAG AGG CTC CAT CGG AAC CTC CTT AAG
T E D A A H R A L L S E R L H R N L L K
1261/421 1291/431
CGC CTC GAG GGG GAG GAG AAG CTC CTT TGG CTC TAC CAC GAG GTG GAA AAG CCC CTC TCC
R L E G E E K L L W L Y H E V E K P L S
1321/441 1351/451
CGG GTC CTG GCC CAC ATG GAG GCC ACC GGG GTA CGG CTG GAC GTG GCC TAC CTT CAG GCC
R V L A H M E A T G V R L D V A Y L Q A
1381/461 1411/471
CTT TCC CTG GAG CTT GCG GAG GAG ATC CGC CGC CTC GAG GAG GAG GTC TTC CGC TTG GCG
L S L E L A E E I R R L E E E V F R L A
1441/481 1471/491
GGC CAC CCC TTC AAC CTC AAC TCC CGG GAC CAG CTG GAA AGG GTG CTC TTT GAC GAG CTT
G H P F N L N S R D Q L E R V L F D E L
1501/501 1531/511
AGG CTT CCC GCC TTG GGG AAG ACG CAA AAG ACA GGC AAG CGC TCC ACC AGC GCC GCG GTG
R L P A L G K T Q K T G K R S T S A A V
1561/521 1591/531
CTG GAG GCC CTA CGG GAG GCC CAC CCC ATC GTG GAG AAG ATC CTC CAG CAC CGG GAG CTC
L E A L R E A H P I V E K I L Q H R E L
1621/541 1651/551
ACC AAG CTC AAG AAC ACC TAC GTG GAC CCC CTC CCA AGC CTC GTC CAC CCG AGG ACG GGC
T K L K N T Y V D P L P S L V H P R T G
1681/561 1711/571
CGC CTC CAC ACC CGC TTC AAC CAG ACG GCC ACG GCC ACG GGG AGG CTT AGT AGC TCC GAC
R L H T R F N Q T A T A T G R L S S S D
1741/581 1771/591
CCC AAC CTG CAG AAC ATC CCC GTC CGC ACC CCC TTG GGC CAG AGG ATC CGC CGG GCC TTC
P N L Q N I P V R T P L G Q R I R R A F
1801/601 1831/611
GTG GCC GAG GCG GGT TGG GCG TTG GTG GCC CTG GAC TAT AGC CAG ATA GAG CTC CGC GTC
V A E A G W A L V A L D Y S Q I E L R V
1861/621 1891/631
CTC GCC CAC CTC TCC GGG GAC GAA AAC CTG ATC AGG GTC TTC CAG GAG GGG AAG GAC ATC
L A H L S G D E N L I R V F Q E G K D I
1921/641 1951/651
CAC ACC CAG ACC GCA AGC TGG ATG TTC GGC GTC CCC CCG GAG GCC GTG GAC CCC CTG ATG
H T Q T A S W M F G V P P E A V D P L M
1981/661 2011/671
CGC CGG GCG GCC AAG ACG GTG AAC TTC GGC GTC CTC TAC GGC ATG TCC GCC CAT AGG CTC
R R A A K T V N F G V L Y G M S A H R L
2041/681 2071/691
TCC CAG GAG CTT GCC ATC CCC TAC GAG GAG GCG GTG GCC TTT ATA GAG CGC TAC TTC CAA
S Q E L A I P Y E E A V A F I E R Y F Q
2101/701 2131/711
AGC TTC CCC AAG GTG CGG GCC TGG ATA GAA AAG ACC CTG GAG GAG GGG AGG AAG CGG GGC
S F P K V R A W I E K T L E E G R K R G
2161/721 2191/731
TAC GTG GAA ACC CTC TTC GGA AGA AGG CGC TAC GTG CCC GAC CTC AAC GCC CGG GTG AAG
Y V E T L F G R R R Y V P D L N A R V K
2221/741 2251/751
AGC GTC AGG GAG GCC GCG GAG CGC ATG GCC TTC AAC ATG CCC GTC CAG GGC ACC GCC GCC
S V R E A A E R M A F N M P V Q G T A A
2281/761 2311/771
GAC CTC ATG AAG CTC GCC ATG GTG AAG CTC TTC CCC CGC CTC CGG GAG ATG GGG GCC CGC
D L M K L A M V K L F P R L R E M G A R
2341/781 2371/791
ATG CTC CTC CAG GTC CAC GAC GAG CTC CTC CTG GAG GCC CCC CAA GCG CGG GCC GAG GAG
M L L Q V H D E L L L E A P Q A R A E E
2401/801 2431/811
GTG GCG GCT TTG GCC AAG GAG GCC ATG GAG AAG GCC TAT CCC CTC GCC GTG CCC CTG GAG
V A A L A K E A M E K A Y P L A V P L E
2461/821 2491/831
GTG GAG GTG GGG ATG GGG GAG GAC TGG CTT TCC GCC AAG GGT TAG
V E V G M G E D W L S A K G *

FIG. 4
(sheet 2)

FIG. 5
(Sheet 1)

DNA and protein sequence of the coding region of pMR8, encoding FY4

1/1 31/11
ATG CTG GAA CGT CTG GAA TTC GGC AGC CTC CTC CAC GAG TTC GGC CTC CTG GAG GCC CCC
M L E R L E F G S L L H E F G L L E A P
61/21 91/31
GCC CCC CTG GAG GAG GCC CCC TGG CCC CCG CCG GAA GGG GCC TTC GTG GGC TTC GTC CTC
A P L E E A P W P P P E G A F V G F V L
121/41 151/51
TCC CGC CCC GAG CCC ATG TGG GCG GAG CTT AAA GCC CTG GCC GCC TGC AGG GAC GGC CGG
S R P E P M W A E L K A L A A C R D G R
181/61 211/71
GTG CAC CGG GCA GCA GAC CCC TTG GCG GGG CTA AAG GAC CTC AAG GAG GTC CGG GGC CTC
V H R A A D P L A G L K D L K E V R G L
241/81 271/91
CTC GCC AAG GAC CTC GCC GTC TTG GCC TCG AGG GAG GGG CTA GAC CTC GTG CCC GGG GAC
L A K D L A V L A S R E G L D L V P G D
301/101 331/111
GAC CCC ATG CTC CTC GCC TAC CTC CTG GAC CCC TCC AAC ACC ACC CCC GAG GGG GTG GCG
D P M L L A Y L L D P S N T T P E G V A
361/121 391/131
CGG CGC TAC GGG GGG GAG TGG ACG GAG GAC GCC GCC CAC CGG GCC CTC CTC TCG GAG AGG
R R Y G G E W T E D A A H R A L L S E R
421/141 451/151
CTC CAT CGG AAC CTC CTT AAG CGC CTC GAG GGG GAG GAG AAG CTC CTT TGG CTC TAC CAC
L H R N L L K R L E G E E K L L W L Y H
481/161 511/171
GAG GTG GAA AAG CCC CTC TCC CGG GTC CTG GCC CAC ATG GAG GCC ACC GGG GTA CGG CTG
E V E K P L S R V L A H M E A T G V R L
541/181 571/191
GAC GTG GCC TAC CTT CAG GCC CTT TCC CTG GAG CTT GCG GAG GAG ATC CGC CGC CTC GAG
D V A Y L Q A L S L E L A E E I R R L E
601/201 631/211
GAG GAG GTC TTC CGC TTG GCG GGC CAC CCC TTC AAC CTC AAC TCC CGG GAC CAG CTG GAA
E E V F R L A G H P F N L N S R D Q L E
661/221 691/231
AGG GTG CTC TTT GAC GAG CTT AGG CTT CCC GCC TTG GGG AAG ACG CAA AAG ACA GGC AAG
R V L F D E L R L P A L G K T Q K T G K
721/241 751/251
CGC TCC ACC AGC GCC GCG GTG CTG GAG GCC CTA CGG GAG GCC CAC CCC ATC GTG GAG AAG
R S T S A A V L E A L R E A H P I V E K
781/261 811/271
ATC CTC CAG CAC CGG GAG CTC ACC AAG CTC AAG AAC ACC TAC GTG GAC CCC CTC CCA AGC
I L Q H R E L T K L K N T Y V D P L P S
841/281 871/291
CTC GTC CAC CCG AGG ACG GGC CGC CTC CAC ACC CGC TTC AAC CAG ACG GCC ACG GCC ACG
L V H P R T G R L H T R F N Q T A T A T
901/301 931/311
GGG AGG CTT AGT AGC TCC GAC CCC AAC CTG CAG AAC ATC CCC GTC CGC ACC CCC TTG GGC
G R L S S S D P N L Q N I P V R T P L G
961/321 991/331
CAG AGG ATC CGC CGG GCC TTC GTG GCC GAG GCG GGT TGG GCG TTG GTG GCC CTG GAC TAT
Q R I R R A F V A E A G W A L V A L D Y
1021/341 1051/351

FIG 5.
(Sheet 2)

AGC CAG ATA GAG CTC CGC GTC CTC GCC CAC CTC TCC GGG GAC GAA AAC CTG ATC AGG GTC
S Q I E L R V L A H L S G D E N L I R V
1081/361 1111/371
TTC CAG GAG GGG AAG GAC ATC CAC ACC CAG ACC GCA AGC TGG ATG TTC GGC GTC CCC CCG
F Q E G K D I H T Q T A S W M F G V P P
1141/381 1171/391
GAG GCC GTG GAC CCC CTG ATG CGC CGG GCG GCC AAG ACG GTG AAC TAC GGC GTC CTC TAC
E A V D P L M R R A A K T V N Y G V L Y
1201/401 1231/411
GGC ATG TCC GCC CAT AGG CTC TCC CAG GAG CTA GCC ATC CCC TAC GAA GAA GCG GTG GCC
G M S A H R L S Q E L A I P Y E E A V A
1261/421 1291/431
TTT ATA GAG CGC TAC TTC CAA AGC TTC CCC AAG GTG CGG GCC TGG ATA GAA AAG ACC CTG
F I E R Y F Q S F P K V R A W I E K T L
1321/441 1351/451
GAG GAG GGG AGG AAG CGG GGC TAC GTG GAA ACC CTC TTC GGA AGA AGG CGC TAC GTG CCC
E E G R K R G Y V E T L F G R R R Y V P
1381/461 1411/471
GAC CTC AAC GCC CGG GTG AAG AGC GTC AGG GAG GCC GCG GAG CGC ATG GCC TTC AAC ATG
D L N A R V K S V R E A A E R M A F N M
1441/481 1471/491
CCC GTC CAG GGC ACC GCC GCC GAC CTC ATG AAG CTC GCC ATG GTG AAG CTC TTC CCC CGC
P V Q G T A A D L M K L A M V K L F P R
1501/501 1531/511
CTC CGG GAG ATG GGG GCC CGC ATG CTC CTC CAG GTC CAC GAC GAG CTC CTC CTG GAG GCC
L R E M G A R M L L Q V H D E L L L E A
1561/521 1591/531
CCC CAA GCG CGG GCC GAG GAG GTG GCG GCT TTG GCC AAG GAG GCC ATG GAG AAG GCC TAT
P Q A R A E E V A A L A K E A M E K A Y
1621/541 1651/551
CCC CTC GCC GTG CCC CTG GAG GTG GAG GTG GGG ATG GGG GAG GAC TGG CTT TCC GCC AAG
P L A V P L E V E V G M G E D W L S A K
1681/561
GGT TAG
G *

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 96 30 3880

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
D,A	WO-A-92 06188 (BARNES WAYNE M) * the whole document *	1-9	C12N15/54 C12N9/12 C12Q1/68
A	WO-A-91 09944 (CETUS CORP) * the whole document *	1-9	
A	WO-A-94 05797 (KISELEV VSEVOLOD ;SEVERIN EVGENII (RU); KORPELA TIMO (FI)) * the whole document *	1-9	
D,A	EUR. J. BIOCHEM. (1992), 209(1), 351-5 CODEN: EJBCAI;ISSN: 0014-2956, 1992, XP000578012 RICHTER, OLIVER MATTHIAS H. ET AL: "Cloning and sequencing of the gene for the cytoplasmic inorganic pyrophosphatase from the thermoacidophilic archaebacterium Thermoplasma acidophilum" * the whole document *	1-9	
A	WO-A-90 12111 (HARVARD COLLEGE) * the whole document *	1-9	TECHNICAL FIELDS SEARCHED (Int.Cl.6)
P,X, D	EP-A-0 655 506 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE, USA) * page 6, line 15 - line 17; claims 1-52 *	1,6,9	C12N C12Q
P,X	NATURE, vol. 376, 31 August 1995, MACMILLAN JOURNALS LTD., LONDON,UK, pages 796-797, XP002009831 M.A. REEVE AND C.W. FULLER: "A novel thermostable polymerase for DNA sequencing" * the whole document *	1-9	

-/--

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	20 September 1996	Hornig, H

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 96 30 3880

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
P,X	AMERSHAM LIFE SCIENCE, EDITORIAL COMMENTS, vol. 22, no. 2, July 1995, pages 29-36, XP002009832 S.B. SAMOLS ET AL.: "Thermo Sequenase; a new thermostable DNA polymerase for DNA sequencing" * the whole document *	1-9	
P,A	PROC. NATL. ACAD. SCI. U. S. A. (1995), 92(14), 6339-43 CODEN: PNASA6;ISSN: 0027-8424, 3 July 1995, XP002009833 TABOR, STANLEY ET AL: "A single residue in DNA polymerases of the Escherichia coli DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides" * the whole document *	1-9	
			TECHNICAL FIELDS SEARCHED (Int.Cl.6)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	20 September 1996	Hornig, H

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)